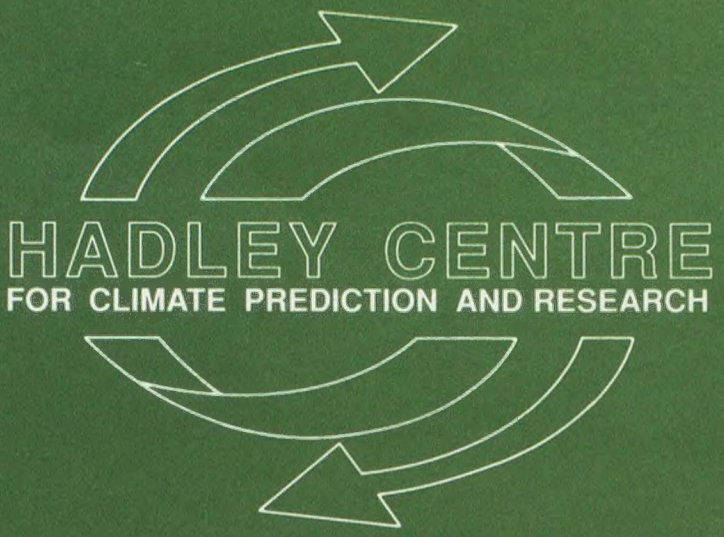


The incidence of extremes in worldwide and Central England temperatures

by

E B Horton, C K Folland and D E Parker

CRTN 83

March 1998

CLIMATE
RESEARCH
TECHNICAL
NOTE

Hadley Centre
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SY

CLIMATE RESEARCH TECHNICAL NOTE NO. 83

THE INCIDENCE OF EXTREMES IN WORLDWIDE AND CENTRAL ENGLAND TEMPERATURES

by

E. B. Horton, C. K. Folland and D. E. Parker

submitted to Climatic Change, December, 1997

Hadley Centre for Climate Prediction and Research
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SY

NOTE: This paper has not been published. Permission to quote from it should be obtained from the Director of the Hadley Centre.

The incidence of extremes in worldwide and Central England temperatures

E.B. Horton, C.K. Folland and D.E. Parker
Hadley Centre
Meteorological Office, Bracknell, UK

Abstract

We describe the development of global fields of 5° latitude × 5° longitude resolution temperature anomalies corresponding to various percentiles. These values are found by fitting anomalies from the 1961-90 base period to 3-parameter gamma distributions. An empirical method is also used to check the validity of the results. Fields of percentiles created using this technique can be used to map the distribution of rare temperature anomalies across the globe for any month, season, year or period of years up to a decade, from 1851 onwards. We focus on the calculation of annual temperature anomaly percentiles. We document and interpret changes in the incidence of extreme annual temperatures since the end of the 19th century. We also apply a similar technique to assess changes in the incidence of extreme daily Central England temperature anomalies.

1. Introduction

For global assessment of extremes, our objective is to calculate, for each 5° latitude × 5° longitude gridbox, the annual or seasonal temperature anomalies which are expected to be reached or exceeded (100-x)% of the time for high percentiles (>50%), or not to be exceeded x% of the time for low percentiles (<50%). The percentiles are chosen as follows: 2%, 5%, 10%, 20%, 30%, 40%, (50%), 60%, 70%, 80%, 90%, 95%, 98%. From these statistics, the rarity of any specified annual anomaly can be assessed, and trends in the worldwide incidence of extremes can be monitored and interpreted. In this paper we present results for the 10th and 90th percentiles. A similar procedure is applied to daily Central England temperatures: for this longer data set we assess changes in extremes using the 5th and 95th percentiles. This paper extends work reported by Jones et al. (1997a) following the Workshop on Indices and Indicators for Climate Extremes held at Asheville, NC, USA, 3-6 JUNE 1997.

2. Data

The gridded global data used for finding the worldwide anomalies at each percentile are the 5° latitude by 5° longitude resolution monthly blended land and sea surface temperatures relative to 1961-90 (Jones (1994), Parker et al. 1995)), over the same period. The monthly fields were averaged into seasonal, then annual fields. Seasonal values were calculated provided at least one

monthly anomaly was available. Annual values were only calculated if all four seasonal values were available. This resulted in 30 incomplete fields on which to base the percentile calculations. Only grid boxes with more than 20 anomalies out of the maximum 30 had their percentile anomalies calculated. Data since the late 19th century from the same data bases were then used for assessments of changes in the incidence of extremes.

For daily Central England temperature, we used an update of the series developed by Parker et al. (1992).

3. Empirical method

In each gridbox, temperature anomalies at each of the chosen percentiles were identified using the method due to Chegodayev (1953) and described fully in Jenkinson (1977). This method provides an estimate of the most probable value of the percentile that is associated with each value in a set of ranked data. A brief description appears in Appendix A. Missing data were only accounted for in that as above, only boxes with more than 20 annual anomalies were used. The values were stored for comparison with those found from fitting the gamma distributions to the anomalies.

4. The 2-parameter gamma distribution

There is no perfect method for fitting a distribution to a set of data with an arbitrary but unimodal distribution. The latter is assumed here. The gamma distribution is however very convenient and can in principle deal with reasonable variations in distribution shape, and with varying skewness. Note that because we are interested in most of the distribution we have not used extreme value methods for the rarest percentiles. We recognise, however, that this is an issue that may be needed to be tackled in the future.

The general form of the 2-parameter gamma density function is given below:

$$f(x) = \frac{1}{b^a \, \Gamma(a)} \int_0^\infty e^{-x/b} \, x^{a-1} \qquad (1)$$

where x is the variable being studied, here temperature anomalies (see section 5 for the problem of negative anomalies), and $\Gamma(a) = (a-1)!$ is the gamma function.

This gamma distribution is fully described by the two parameters a and b. These can be estimated as a' and b' most simply using the method of

moments:

$$a'=\mu^2/\sigma^2 \qquad b'=\sigma^2/\mu. \qquad (2a)$$

where μ = mean of the sample and σ = standard deviation of the sample (Walpole and Myers, 1978).

The frequency distribution curve is positive definite. Changing the values of a and b changes the shape and scale of the gamma curve (Figure 1). Parameter a has a strong influence on the shape of the curve, because, for example, a small value of a means that μ is small (compared with σ), causing a sharp peak in the curve because it is constrained to be positive. Parameter b has a stronger influence on the width of the curve via σ. The distribution is always positively skewed, with a coefficient of skewness, $G=2a^{-1/2}$ (Haktanir, 1991). As the value of a increases, both the skewness and kurtosis decrease, and the distribution becomes more Gaussian. Ropelewski and Jalickee (1983) state that for practical purposes the differences between the gamma and normal distributions is negligible for values of a greater than 50. This is confirmed by Figure 2. With a=51, the coefficient of skewness, G, is 0.28.

For highly skewed distributions, with low values of a, the method of moments (Eq. 2a) becomes less reliable because it does not use all of the information given in the sample (Haan, 1977). This problem becomes significant when the value of a is less than 10 (Haan, 1977). Therefore we choose a maximum likelihood method to estimate the parameters of the gamma distribution, irrespective of the value of a. This method mirrors the principle of the Chegodayev method (Thom, 1957; Essenwanger, 1985) because it will estimate the most probable percentile value corresponding to specific values in the distribution. The maximum likelihood estimates of a and b are calculated from the following:

$$a=(1+\sqrt{1+4A/3})/4A \qquad (2b)$$

with

$$A=\ln\mu - \frac{1}{N}\Sigma \ln x_i \qquad (2c)$$

where x_i = ith temperature anomaly
μ = sample mean
N = sample size

Here we drop the superscripts on a and b when discussing maximum likelihood estimates.

5. The 3-parameter gamma distribution

To fit the gamma distribution to anomalies for the period 1961-90, temperature anomaly distributions (which may be positively or negatively skewed) must be adjusted to make them positive definite, because gamma distributions are only defined for x>0. The possibility of extreme negative anomalies must be taken into account. Our solution to these problems was first to check the sign of the skewness in the untransformed anomaly, then to introduce a location parameter c, and transform the data using one of the following equations:

For data with positive skewness:

$$z_i = \frac{x_i - \mu_x}{\sigma_x} + c \qquad (3a)$$

with

$$c = q + |\min(x)| \qquad (3b)$$

For data with negative skewness:

$$z_i = \frac{\mu_x - x_i}{\sigma_x} + c \qquad (4a)$$

with

$$c = q + |\max(x)| \qquad (4b)$$

where min(x) and max(x) are the minimum and maximum **normalised** values in the series in 1961-90 respectively and q is an arbitrary number of standard deviations discussed below. These transformations accommodate cases where an anomaly is up to q standard deviations more negative/positive than the minimum/maximum anomaly during the base period (1961-90), depending upon the skewness of the data. The kurtoses of the distributions x and z are unchanged by either of the above transformations. The magnitude of the coefficient of skewness is also unchanged, but in Equations 4a and b the sign of the skewness has been changed from negative to positive.

It is now possible, by arbitrarily assigning a value to q, to calculate a and b and so fit a gamma distribution to each transformed distribution. Each grid box and calendar period (e.g. monthly) temperature will have different values for a and b. The variance of the transformed distribution z is always 1, and the mean of z is c.

In most cases there is little difference between a calculated using the maximum

likelihood method and a' calculated from the method of moments. $a'=\mu^2/\sigma^2$ so that with $\sigma^2_z=1$ and $\mu_z=c$, a is approximately equal to c^2. The value of c depends on that assigned to q, so q controls the shape of the fitted gamma distribution. Increasing q causes the fitted distribution to become less skew. To ensure that the skewness of the fitted distribution is as similar as possible to that of the observed distribution, an iterative method was used to determine the best fit value of q. A first guess minimum value was q = 0.5. The parameters a and b, and the skewness of the fitted distribution were then calculated. q was then increased repeatedly by 0.5 until the difference between the observed and fitted skewness reached a minimum, whereupon the value of q used was that which gave the closest fit to the skewness. If the value calculated for a exceeded 50, the distribution was not significantly different to the Gaussian (Ropelewski and Jalickee, 1983), so q was assigned the value used in the previous iteration.

6. Testing for goodness of fit of worldwide 5° latitude by 5° longitude gridded annual data

6.1. *Comparison of skewness*

A comparison between the skewnesses of the fitted gamma distributions and the observed distributions shows that of the 1642 fitted annual distributions, 992 were within 0.1, 1307 were within 0.2 and 1635 were within 0.3 of the skewness of the observed distributions.

6.2. *Cumulative frequency distribution plots*

A visual check was made on the fit by plotting the cumulative frequency distribution of the observed and fitted annual data. To calculate the cumulative frequency distribution of the gamma fits the value of $\Gamma(a)$ must be found. A single precision function in the Wave-Advantage package version 4.2 (Visual Numerics Inc., 1993) calculates $\Gamma(a)$, provided $a<34.5$. For $34.5<a<50$ a look-up table was calculated using a FORTRAN77 NAG routine (Numerical Algorithms Group Ltd., 1993). The iterative method for determining the best fit ensured that there were no instances where a exceeded 50. Cumulative frequency distributions for 4 chosen gridboxes are given in Figure 3 and discussed later.

6.3. *Chi-squared tests of observed and fitted distributions*

A χ^2 test using 4 classes was performed at each grid box to test the null hypothesis that the gamma distribution determined by a and b and the transformed distribution z were from the same population. Four is the maximum number of classes allowed if the χ^2 test is to avoid having fewer than 5 observations in any class, because the minimum number of observations in any time series is defined as 21.

Out of 1642 boxes, 1604 passed the χ^2 test at the 10% significance level. Figure 3 compares the cumulative frequency distributions for a selection of grid boxes with their fitted gamma and Gaussian distributions for annual

temperature anomalies. The cumulative frequency distributions for the boxes best and worst fitted by gamma distributions, which were at 47.5°N, 102.5°E and 12.5°S, 37.5°E respectively are given in Figures 3a and b. Even in the worst fit case, the gamma distribution is no poorer than the Gaussian. Figures 3c and d show the cumulative frequency distributions with the greatest and least skewness, at 2.5°S, 82.5°W and 27.5°S, 107.5°E respectively. The gamma distribution is much better than the Gaussian in the skew example (Figure 3c). Of the 1642 anomaly distributions, 794 have skewness > 0.28, i.e. a < 50. The gamma fit is therefore probably an improvement over the Gaussian in these cases (Section 4).

The χ^2 test was repeated, but this time to test the null hypothesis that the Gaussian distribution determined by the mean and standard deviation and the observed distribution were from the same population. 1604 boxes passed the normality test at the 10% significance level, the same number that passed the gamma fit test. However, we repeated both tests using data for January and found that 1661 boxes passed the gamma test, but only 1310 boxes passed the normality test. We chose January because land surface air temperature distributions are most skewed in winter.

6.4. *Discussion*

The nature of temperature anomalies is such that it is probably impossible to fit one type of distribution curve to the data in all of the gridboxes. The gamma distribution was chosen because of its flexibility in allowing for the expected skewness in monthly or seasonal temperature anomalies, particularly those from the interiors of large land masses. The χ^2 test revealed that in the case of annual mean temperature anomalies, use of the Gaussian distribution would have been just as valid. However, Figures 3c and d show that for the more skew distributions the gamma fit is better, while for the less skew distributions the gamma fit is very similar to that of the Gaussian. It is considered best to use the gamma distribution on all of the scales for consistency.

7. Calculation of anomalies at the chosen percentiles

Once the shape of the distribution has been determined for each box, by calculation of the parameters a, b and q, the values of the gamma distribution must be found at each of the chosen percentiles and transformed back to degrees Celsius using the inverse of equations 3a/b or 4a/b as appropriate.

The Wave-Advantage library (Visual Numerics, Inc., 1993) contains a function which, given a value z and a shape parameter a, returns the probability that a random value from that distribution takes a value less than z. Here z is measured in the transformed (gamma) coordinates and the parameter b is assumed to be 1. However the inverse of this function is needed to find the value $z_{(p)}$ corresponding to a specified probability p that a random value from the distribution takes a value less than $z_{(p)}$. There are several methods of

estimating $z_{(p)}$ which utilise the more accurate algorithms designed to calculated the value of the Gaussian distribution at percentile p. So we used the Wilson-Hilferty approximation which is widely used in flood frequency analysis (Haktanir, 1991):

$$z_{(p)}=a\ [\ 1-\frac{1}{9a}\ \pm\ X_{(p)}(\frac{1}{9a})^{1/2}\ ]^3 \qquad (5)$$

where a is the shape parameter defining the gamma distribution function.
$X_{(p)}$ is the value of the Gaussian distribution function at percentile p.

In this way, once a and b have been determined for the distribution, then the values at each of the chosen percentiles can be found. For distributions where $b\neq1$, the values of $z_{(p)}$ returned must be multiplied by b before being used in the transformation back into anomalies in °C. This is because $\mu=a'b'$. Two examples are given in Figure 4 illustrating the effect on the values of $z_{(p')} = bz_{(p)}$ at the 2nd percentile of changing b. The full transformations are as follows:

When the skewness of the original distribution is positive:

$$x_{(p)} = \mu_x+\sigma_x(b.z_{(p)}-c) \qquad (6a)$$

When the skewness of the original distribution is negative:

$$x_{(p)} = \mu_x-\sigma_x(b.z_{(p)}-c) \qquad (6b)$$

where $z_{(p)}$ = the value corresponding to the pth percentile
$x_{(p)}$ = the temperature anomaly in °C at the pth percentile

Haktanir (1991) established that for most practical purposes the Wilson-Hilferty approximation is sufficiently accurate for distributions with skewness G in the range $-1<G<1$. This includes most of the annual temperature anomaly distributions. The Wilson-Hilferty approximation becomes more accurate as the data become less skewed because it is based on the Gaussian distribution.

The most skewed distribution in our data had a skewness coefficient of 2 with a=1. However, tests based on tabulations by Haktanir (1991) showed that the Wilson-Hilferty approximation yielded a maximum error of 0.023 in $z_{(p)}$ giving a temperature error of $\sigma bz_{(p)}$ =0.03°C. Since temperature anomalies are only accurate to the first decimal place the Wilson-Hilferty approximation is sufficiently accurate for this application.

8. Analyses and results: worldwide gridded data

8.1. *Analyses of particular years.*

The temperature anomalies at the chosen percentiles were calculated for the period 1961-90 using both Chegoyadev's empirical method and the gamma method. Annual anomalies for 1995 (the warmest year globally to that date) and 1964 (a relatively cool year) were calculated and each gridbox assigned a percentile from each method. Charts showing the allotted percentiles for the gamma and empirical methods are given in Figures 5a, c and b, d respectively. The charts for 1964 (Figure 5a,b) are in good agreement over most of the world. Only isolated locations in the North Atlantic, North Pacific and Amazonia were as warm as the 98% value, whereas large areas of the subtropical South Atlantic and southwest Indian Ocean, and part of the Pacific west of Peru were colder than the 2% value. The 1995 charts (Figure 5c,d) are also in good agreement, with concentrated areas of unusual warmth west of the Iberian Peninsula and over the former Soviet Union. Other areas of unusual warmth were around Indonesia and Papua-New Guinea, southeastern Africa (Mozambique, Tanzania and Zambia) and on the west coast of North America. The maximum likelihood gamma method is to be preferred because in principle it will give more consistent results for the far extremes. The very good agreement between the two methods, however, is a strong indication that the gamma method is working well.

Cumulative frequency distribution curves based on the gamma percentiles were also plotted for 1964 and 1995. These are given in Figures 6a and b along with the cumulative frequency distribution curve for all years from 1961-90. When the solid curve lies above the dashed line, more boxes than usual have anomalies *not exceeding* the various percentiles, i.e. there is a preponderance of colder than usual conditions. For example, Figure 6a shows that 65% of boxes in 1964 were colder than their respective 1961-90 medians. When the solid line lies below the dashed line, fewer boxes have anomalies *not exceeding* the percentiles, i.e. there is a preponderance of warmer than usual conditions. An extreme example is in 1995 (Figure 6b), where 85% of gridboxes were warmer than the median. A Kolmogorov-Smirnov test (Siegel and Castellan, 1988) was used to determine whether there was a statistically significant difference between the climatological cumulative distribution and that for each of these two years. The Kolmogorov-Smirnov tests revealed that although 1964 was colder than normal, it was not significantly cold, even at the 10% significance level. Note, however, that the significance may have been slightly underestimated because the cold 1964 data are also in the reference period. When applied to the 1995 data, the test showed that 1995 was significantly warm at the 1% significance level.

8.2. *Trends in worldwide incidence of annual and seasonal temperature extremes.*

Using cumulative frequency distributions like Figure 6, we constructed time-series of the fraction of the monitored global area having extreme temperatures on annual, seasonal and monthly timescales. Trends in these fractional areas

were then estimated using Sen's method (Gilbert 1987), namely by calculating the median slope between all possible pairs of values. This test was chosen over simple linear regression because it is less affected by outliers. The Mann-Kendall rank test (Gilbert 1987) was used to determine whether each trend was significant, using a one-tailed test with H_0: no trend and either H_1: increasing trend (for fractional area with temperatures exceeding 90th percentile) or H_1: decreasing trend (for fractional area with temperatures not exceeding 10th percentile).

Figure 7 shows the percentage of the monitored global area with annual temperature anomalies less than the 10th or exceeding the 90th percentile for 1900-96, as well as the percentage of global surface area with data. There was a strong decline in the proportion of the globe with unusually cold conditions, from about 40% down to 15%, between 1900 and 1940. During this period the area of data coverage doubled, so the proportional decrease in low temperatures could be partly attributed to the increase in total area with data. However the proportion of the monitored area with unusually warm temperature anomalies increased only moderately from 1900 to 1940, indicating that changes in data coverage do not dominate the results. However, during the Second World War, when data were particularly sparse, the area with unusually high temperatures reached a peak, which should be regarded with caution. After 1980 there was a clear rise in the fractional area with temperature anomalies exceeding the 90th percentile and a corresponding drop in the fractional area with temperature anomalies lower than the 10th percentile, to very low values in some years. It is striking how the overall decline in the area of very cold conditions since 1900 is more marked than the increase in the area of very warm conditions.

Figure 8 shows the fractional areas with unusually high and low temperatures for each season. All of the months (not shown) and seasons show a similar pattern to the annual plots, with large fractional areas of low temperatures at the beginning of the period and increasing (decreasing) warm (cold) fractional areas after 1980. All show the striking larger overall decline in very cold conditions than increase in very warm conditions.

The annual and seasonal linear trends (estimated using Sen's method) of the areas covered by areas warmer than the 90th and cooler than the 10th percentile are given in Table 1. All of the trends are statistically significant at the 95% confidence level according to the Mann-Kendall test. Thus although the reduction in very cold areas is more prominent than the increase in very warm areas, both contribute significantly to the observed warming over the last century.

8.3. *Interpretation*

The decline in the incidence of extremely low temperatures exceeded the increase in the incidence in extremely high temperatures (Figures 7 and 8). A straightforward explanation of this is that the negative average global temperature anomaly at the beginning of the 20th century was larger than the

recent average positive anomaly, relative to the 1961-1990 base period (Nicholls et al., 1996). We therefore sought to remove this bias by referencing each year's data to the global average for that year, instead of to the base period. Because land and sea surface temperature anomalies are likely to come from different statistical populations, we subtracted the annual globally-averaged land air temperature anomaly from each gridded land air temperature anomaly in the relevant year, and correspondingly for sea surface temperature. We also recalculated the percentiles from 1961-1990 data treated in this way, so as to remove the effects of all interannual and longer-term variations of the global average over land or sea from the analysis. The residuals should provide an assessment of the incidence of extremes relative to the annual global-average land or oceanic temperature anomaly.

The results for land (Figure 9a) suggest declines in the incidence of both cold and warm extremes: the greater decline is in the latter. An overall decline in the incidence of extremes may be expected as a result of the increasing density of the network of observing stations within gridboxes, reducing the variance of the gridbox anomalies (Jones et al., 1997b). The data were, however, more positively skewed (or less negatively skewed) before 1900 than subsequently. This may reflect the influence of early instrumental exposures: for example, extra overheating of Glaisher and similar thermometer shelters in locally sunnier and warmer than normal springs and summers (Parker, 1994) would encourage overall positive skewness. The oceanic results (Figure 9b) merely suggest greater variance of the gridbox anomalies until the 1950s, again an expected consequence of smaller within-box data samples. Note the increased incidence of both warm and cold oceanic extremes in 1983 and 1987, recent El Niño years. This can be explained in terms of the cold midlatitude Pacific which accompanies El Niño events (Allan et al., 1996).

9. Analyses and results: daily Central England temperature

Jones et al. (1997a) found no long-term change in the incidence of daily Central England temperatures above the 90th percentile, though there was an increase in the recent, warm decade. On the other hand, they found more days below the 10th percentile before 1930 and especially before 1900. Here, using the same smoothed reference annual cycle as Jones et al (1997a) to estimate the anomalies, we repeat the analysis for the 5th and 95th percentiles (Figure 10). Again, the incidence of very cold days is greater than at present before about 1930 (Figure 10a). If annual or monthly anomalies are removed from the data and from the calculation of the reference-period gamma distributions (Figures 10b and c), the decline is reduced but not eliminated, and excess very cold days persist until the 1950s. There is perhaps a slight deficit of very warm days before about 1910 (Figure 10d) but compensation for annual or monthly anomalies generates a slight excess until the 1950s (Figures 10e,f). Consideration of Figures 10b,c and e,f suggests that the *within-year* and *within-month* variance of daily Central England temperatures (excluding that due to the annual cycle) was enhanced until the 1950s, without much change in

skewness. So the incidence of very cold (<5th percentile) and very warm (>95th percentile) days, relative to individual years' or months' averages, was in each case about 6% (perhaps a little more for very cold days) rather than the expected 5%. The data are complete, so data paucity is not an explanation. The expected change in variance resulting from the change from one station to an average of three in 1878 was compensated for by Parker et al. (1992) and has no evident effect on Figures 10c,d and e,f. The replacement of Stonyhurst by the average of Squires Gate and Ringway at the beginning of 1959 coincides approximately with the reduction in within-year and within-month variance in Figures 10b,c and e,f, but according to Parker et al. (1992), this change did not affect the variance of the series. So the decline in within-month variability of daily Central England temperature is likely to be real, and in future work daily Central England temperature and its variability will be investigated in more detail, in the light of atmospheric circulation variations.

10. Conclusions

The 3-parameter gamma distribution is a promising method of analysing temperature extremes. It allows regions with unusual temperature anomalies to be mapped (Figure 5). The reduction, globally, in very cold areas since 1900 is more prominent than the increase in very warm areas on all time-scales from a month to a year. However, compensation for land-specific and ocean-specific global annual average temperature anomalies removes this characteristic, because prevalent global-average negative anomalies in the early 20th century were larger than the recent dominant global-average positive anomalies. The residuals suggest that the earlier data were more variant over both land and ocean, possibly owing to smaller within-gridbox data samples, and the land data were more positively skew, possibly owing to poorer instrumental exposures. The decline in within-year and within-month variability of daily Central England temperature merits further investigation.

The fact that the areas of missing data change strongly through time is the main uncertainty affecting the global results. It will soon be possible to assess this problem as attempts to create unbiased globally complete datasets are being made. However, it is probably also necessary to carry out Monte Carlo simulations to asses this problem and to estimate the size of the uncertainty that time-varying data coverage introduces.

References

Allan,R, Lindesay,J. and Parker,D.E. 1996, *El Niño Southern Oscillation and Climatic Variability (Atlas).* CSIRO Publishing, p. 416.

Chegodayev, N.N., 1953, 'Computation of runoff on small catchments', *Transzhedorizdat*, 1953 (in Russian).

Essenwanger, O.M., 1985, '*World Survey of Climatology: General Survey of Climatology, Volume 1B: Elements of Statistical Analysis*', edited by H.E. Landsberg, Elsevier Science Publishers B.V., Amsterdam, p. 424.

Gilbert, R.O., 1987, '*Statistical methods for environmental pollution monitoring*', Van Nostrand Reinhold Company, New York, p. 384.

Haan, C.T., 1977 '*Statistical methods in hydrology*', Iowa State University Press, Ames, Iowa, p. 378.

Haktanir, T., 1991, 'Practical computation of gamma frequency factors', *Hydrological Sciences Journal,* **36**, 599-610.

Harter, H.L., 1976, 'New tables for percentage points of the Pearson Type III distribution', *Technical Release* **38**, Central Technical Unit, US Soil Conservation Service, USGS, USA.

Jenkinson, A.F., 1977, "The analysis of meteorological and other geophysical extremes', *Synoptic Climatology Branch Memo.* **58**, Meteorological Office, Bracknell, UK, p. 41.

Jones, P.D., 1994, 'Hemispheric surface air temperature variations: a reanalysis and an update to 1993', *J. Climate*, **7**, 1794-1802

Jones, P.D., Horton, E.B., Folland, C.K., Hulme, M., Parker, D.E. and Basnett, T.A., 1997a, 'Climatic extremes: approaches to creating indices from daily and monthly data', Submitted to *Climatic Change*, 1997.

Jones, P.D., Osborn, T.J. and Briffa, K.R., 1997b, 'Estimating errors in large-scale temperature averages', *J. Climate*, **10**, 2548-2568.

Nicholls, N., Gruza, G.V., Jouzel, J., Karl, T.R., Ogallo, L.A. and Parker, D.E., 1996, 'Observed climate variability and change', In '*Climate Change 1995 - The Science of Climate Change'. The second assessment report of the Intergovernmental Panel on Climate Change: Contribution of Working Group I.* J.T. Houghton, L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg and K. Maskell (Eds.). Cambridge University Press, Cambridge, UK, 133-192.

Numerical Algorithms Group Ltd., 1993, '*NAG FORTRAN Library Mark 16*', Oxford, UK.

Parker,D.E., 1994, 'Effects of changing exposure of thermometers at land stations', *Int.J.Climatol.*, **14**, 1-31.

Parker, D.E., C.K. Folland and M. Jackson, 1995, 'Marine surface temperature: observed variations and data requirements', *Climate Change*, **31**, 559-600

Parker, D.E., Legg, T.P. and Folland, C.K., 1992, 'A new daily Central England

Temperature series, 1772-1991', *Int.J.Climatol.*, **12**, 317-342.

Ropelewski, C.F. and J.B. Jalickee, 1983, 'Estimating the significance of seasonal precipitation amounts using approximations of the inverse gamma function over an extended range', *Preprints, 8th conference on probability and statistics in atmospheric sciences, 16-18 Nov 1983, Hot Springs*, Amer. Meteorol. Soc., 125-129

Siegel, S. and Castellan, N.J., 1988, '*Nonparametric statistics for the behavioral sciences*', 2nd Edition, Mc. Graw- Hill Kogakusha Ltd, Tokyo, p. 399.

Thom, H.C.S., 1957 ' A statistical method of evaluating augmentation of precipitation by cloud seeding.' In H.T. Orville (Chairman), *Final Report of the advisory committee on weather control.* Govt. Printing Office, Washington, D.C. 5-25

Walpole, R.E., and R.H. Myers, 1978, '*Probability and statistics for engineers and scientists.*', Macmillan Publishing Co., New York, p. 580.

Visual Numerics, Inc., 1993, '*PV-WAVE Advantage: PV-WAVE Command Language*', Vol. 1: p. 652.; Vol. 2, p. 658., Houston, Texas, USA.

APPENDIX A

Empirical Method of Calculating Percentile Values

Rank a series of N values in ascending order X_1 X_2 ... X_N. The probability that a random value takes a value of less than X_m is

$$p = \frac{m-0.31}{N+0.38}$$

This equation is rearranged in order to find the rank m of the value corresponding to a specific chosen probability p.

$$m = (p(N+0.38))+0.31$$

Note that the values of m corresponding to chosen percentiles (e.g. 10%) are in general not integers.

For the 30 year base period, with all 30 values present the rank of the 98th percentile is 30.08. If only 21 values are present (the minimum number of observations allowed in this study) the rank of the 98th percentile is 21.26. In both cases it is impossible to interpolate above the maximum value in the series, so an approximation was made to arbitrarily attribute the maximum value in the series to the 98th percentile. A similar problem occurs at the bottom of the series, so the minimum value in the series was assigned to the 2nd percentile. Thus the values at the 2nd percentile are likely to be too high, while those at the 98th percentiles are likely to be too low.

Jenkinson (pers. comm.) states that this algorithm works well with all underlying distributions that are not greatly different from Gaussian. It is sometimes used in extreme value analysis where each value being analysed is the biggest or smallest in a given year. Other algorithms exist based on different assumptions about the underlying (extreme value) distribution, but are not greatly different.

Figure Captions

Figure 1. Gamma frequency distribution curves with (top) b=1 and various values of parameter a; and (bottom) a=2 and various values of parameter b.

Figure 2. Comparison of gamma and Gaussian cumulative frequency distributions. The gamma curve is described by parameters a=50, b=0.14. The Gaussian curve has μ=7.07 and σ=1.0.

Figure 3. Examples of fitting gamma and Gaussian distributions to annual temperature anomalies.
a, The distribution best fitted by a gamma cumulative frequency distribution.
b, The distribution worst fitted by a gamma cumulative frequency distribution.
c, The most skew observed distribution.
d, The least skew observed distribution.

Figure 4. The influence of b on the value of the transformed gamma coordinate $z_{(p')} = bz_{(p)}$ corresponding to a particular percentile. The left and right hand graphs are identical, but the abscissa has been marked in percentile points on the left graph and as actual values of $z_{(p')}$ on the right. The top graphs have b=1.5, the bottom graphs have b=2. A random value has 2% chance of being in the shaded region.

Figures 5a,b. Temperature anomalies for 1964, expressed as percentiles of 1961-90 data. Percentile values were calculated in (a, top) by fitting the observations to the gamma distribution and in (b, bottom) by the empirical method of Chegodayev (1953).

Figures 5c,d. Temperature anomalies for 1995, expressed as percentiles of 1961-90 data. Percentile values were calculated in (c, top) by fitting the observations to the gamma distribution and in (d, bottom) by the empirical method of Chegodayev (1953).

Figure 6. Percentages of anomalies which did not exceed the specified anomaly at each percentile during a, 1964, b,1995. Each of these cumulative distributions (solid lines) is compared with that for all years from 1961 to 1990 (dashed lines).

Figure 7. Fraction of the global data area with annual surface temperatures above the 90th percentile (solid line) or below the 10th percentile (dotted line). Also shown, in the inset, is the percentage of global surface area with data. The area with temperature anomalies exceeding the 90th percentile has increased by 0.11% per year, significant at the 95% confidence level. The area with temperature anomalies not exceeding the 10th percentile has decreased by 0.40% per year, also significant at the 95% confidence level.

Figure 8. Fraction of global data area with seasonal surface temperatures above the 90th percentile (solid line) or below the 10th percentile (dashed line). Also shown, in the insets, are the percentages of global surface area with data in each season.

Figure 9a. Fraction of global **land** data area with **adjusted** (see text) annual surface air temperatures above the 90th percentile (solid line) or below the 10th percentile (dotted line). Gridbox anomalies were first adjusted by subtracting the annual globally-averaged land air temperature anomaly in the relevant year. The percentiles were also calculated from 1961-1990 data treated in this way, so as to remove the effects of all interannual and longer-term variations of the global average.

Figure 9b. Fraction of global **ocean** data area with **adjusted** annual sea surface temperatures above the 90th percentile (solid line) or below the 10th percentile (dotted line). Gridbox anomalies were first adjusted by subtracting the annual globally-averaged sea surface temperature anomaly in the relevant year. The percentiles were also calculated from 1961-1990 data treated in this way.

Figure 10a. Fraction of days with Central England temperature not exceeding the 1961-1990 5th percentile, 1772-1996.

Figure 10b. As Figure 10a, but percentiles and numbers of extreme days were calculated using daily anomalies which had been adjusted by subtracting the annual mean anomaly.

Figure 10c. As Figure 10a, but percentiles and numbers of extreme days were calculated using daily anomalies which had been adjusted by subtracting the monthly mean anomaly.

Figure 10d. Fraction of days with Central England temperature exceeding the 1961-1990 95th percentile, 1772-1996.

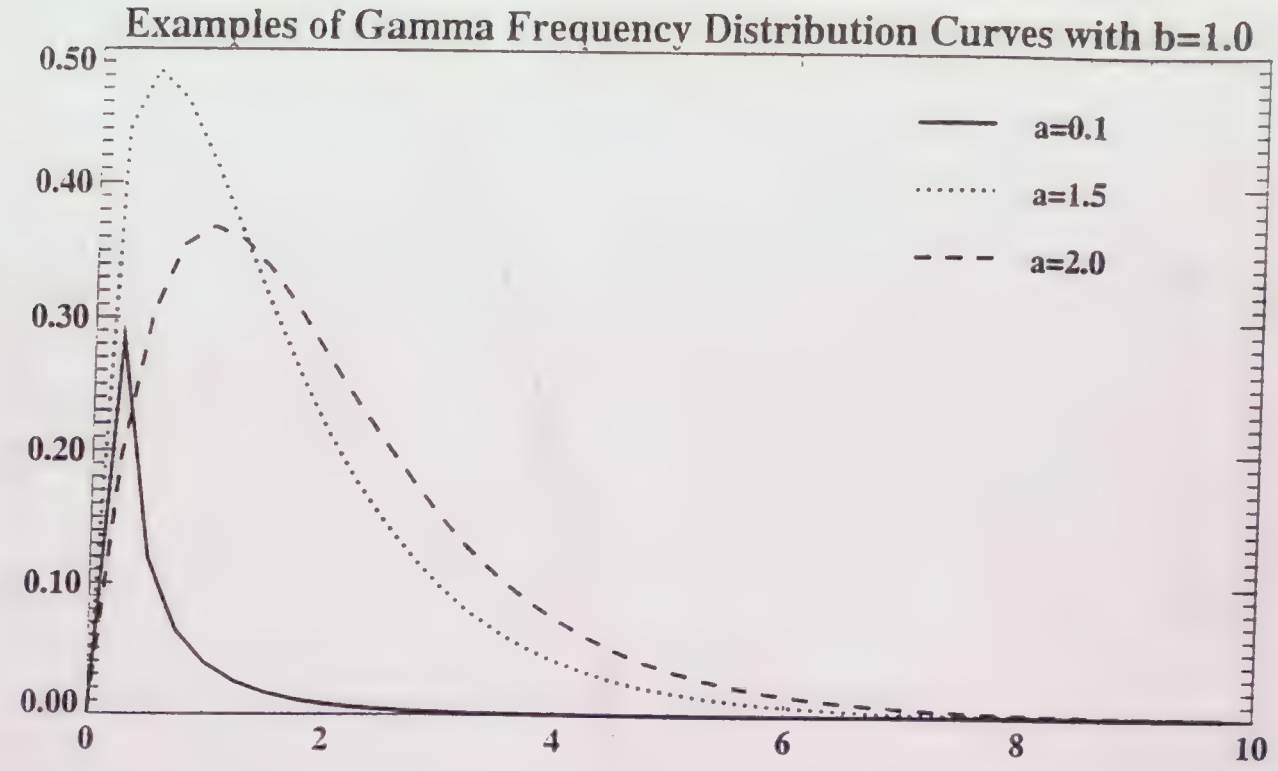
Figure 10e. As Figure 10d, but percentiles and numbers of extreme days were calculated using daily anomalies which had been adjusted by subtracting the annual mean anomaly.

Figure 10f. As Figure 10d, but percentiles and numbers of extreme days were calculated using daily anomalies which had been adjusted by subtracting the monthly mean anomaly.

Table 1. Seasonal and annual trends in the areas with unusually warm or cold temperatures

	Trend in fraction of data area with temperature anomalies exceeding the 90th percentile (% per year)	Trend in fraction of data area with temperature anomalies lower than the 2nd percentile (% per year)
Annual	0.12	-0.45
DJF	0.09	-0.27
MAM	0.09	-0.28
JJA	0.07	-0.30
SON	0.07	-0.32

Table 1. Seasonal and annual trends in the areas with unusually warm or cold temperatures



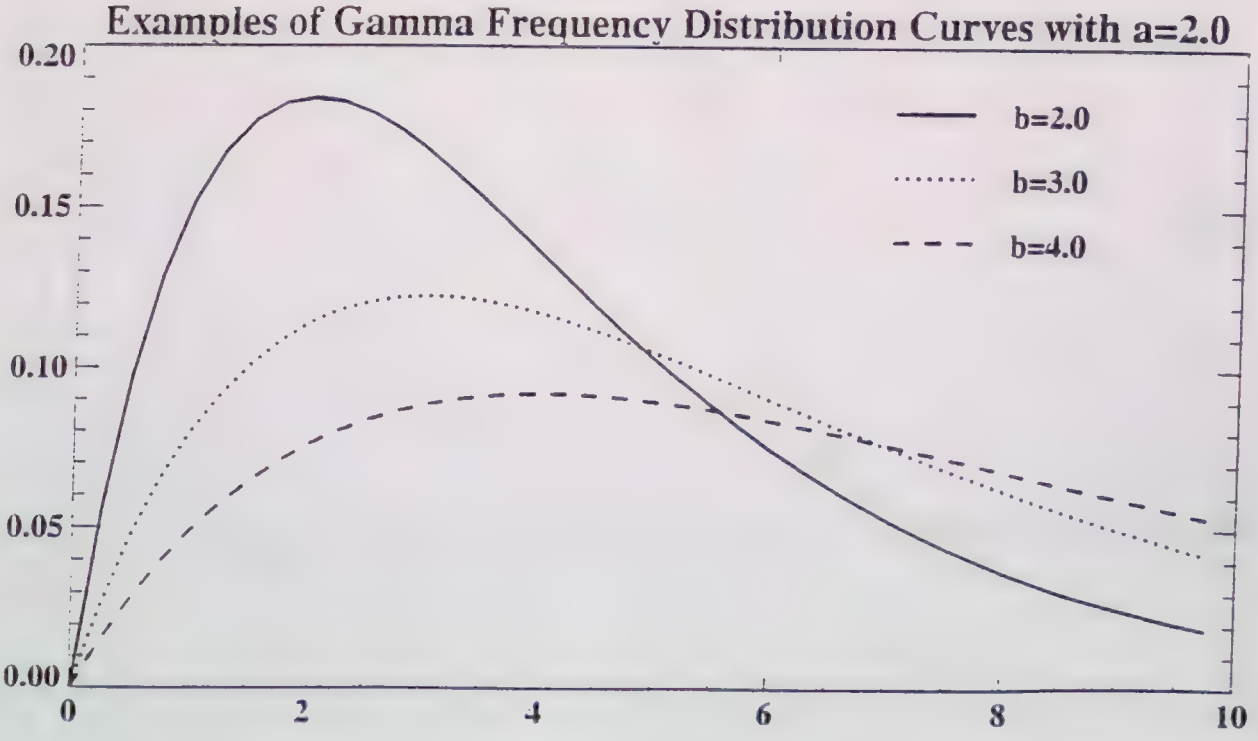


Figure 1. Gamma frequency distribution curves with (top) b=1 and various values of parameter a; and (bottom) a=2 and various values of parameter b.

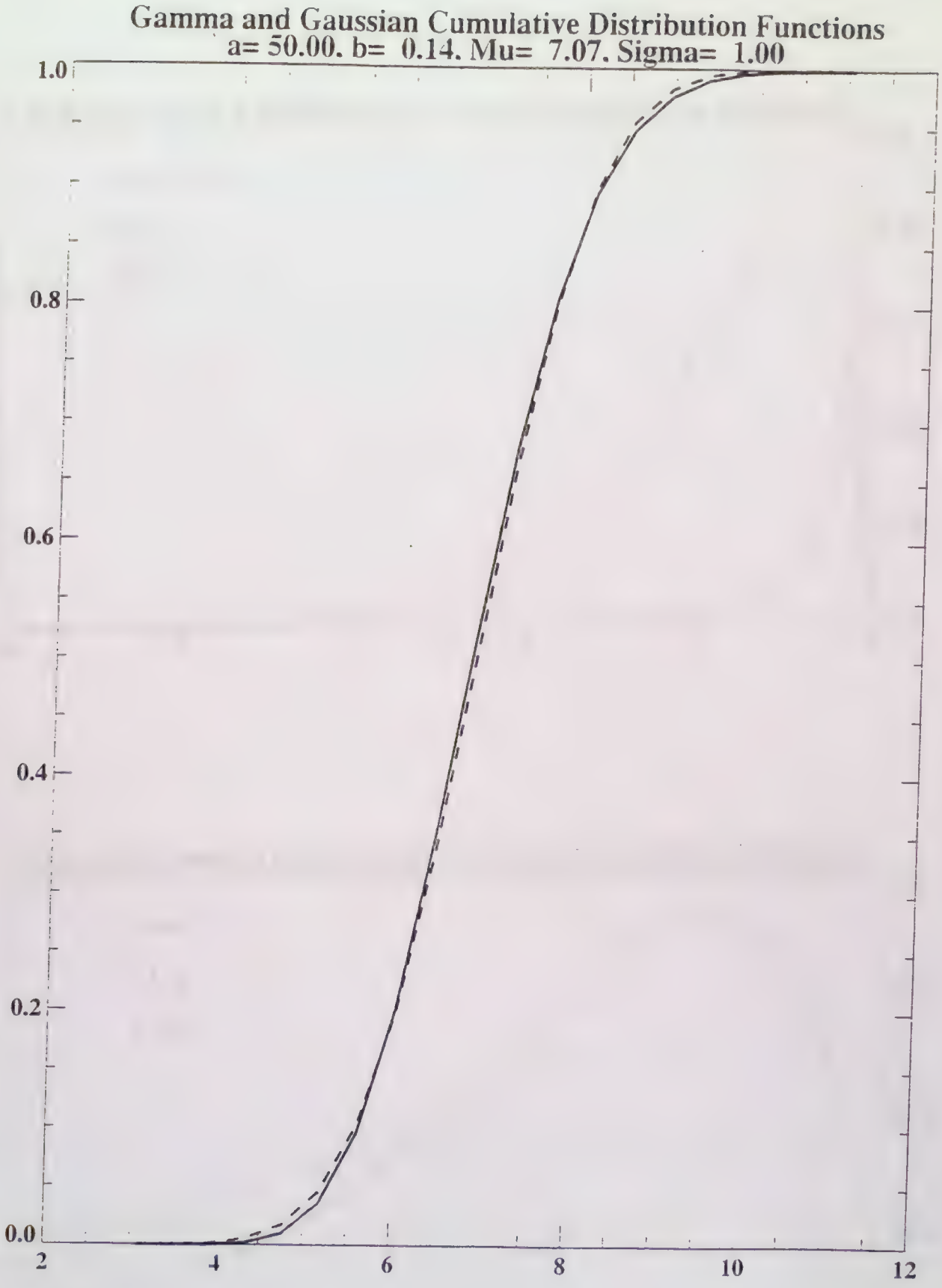


Figure 2. Comparison of gamma and Gaussian cumulative frequency distributions. The gamma curve is described by parameters a=50. b=0.14. The Gaussian curve has μ=7.07 and σ=1.0.

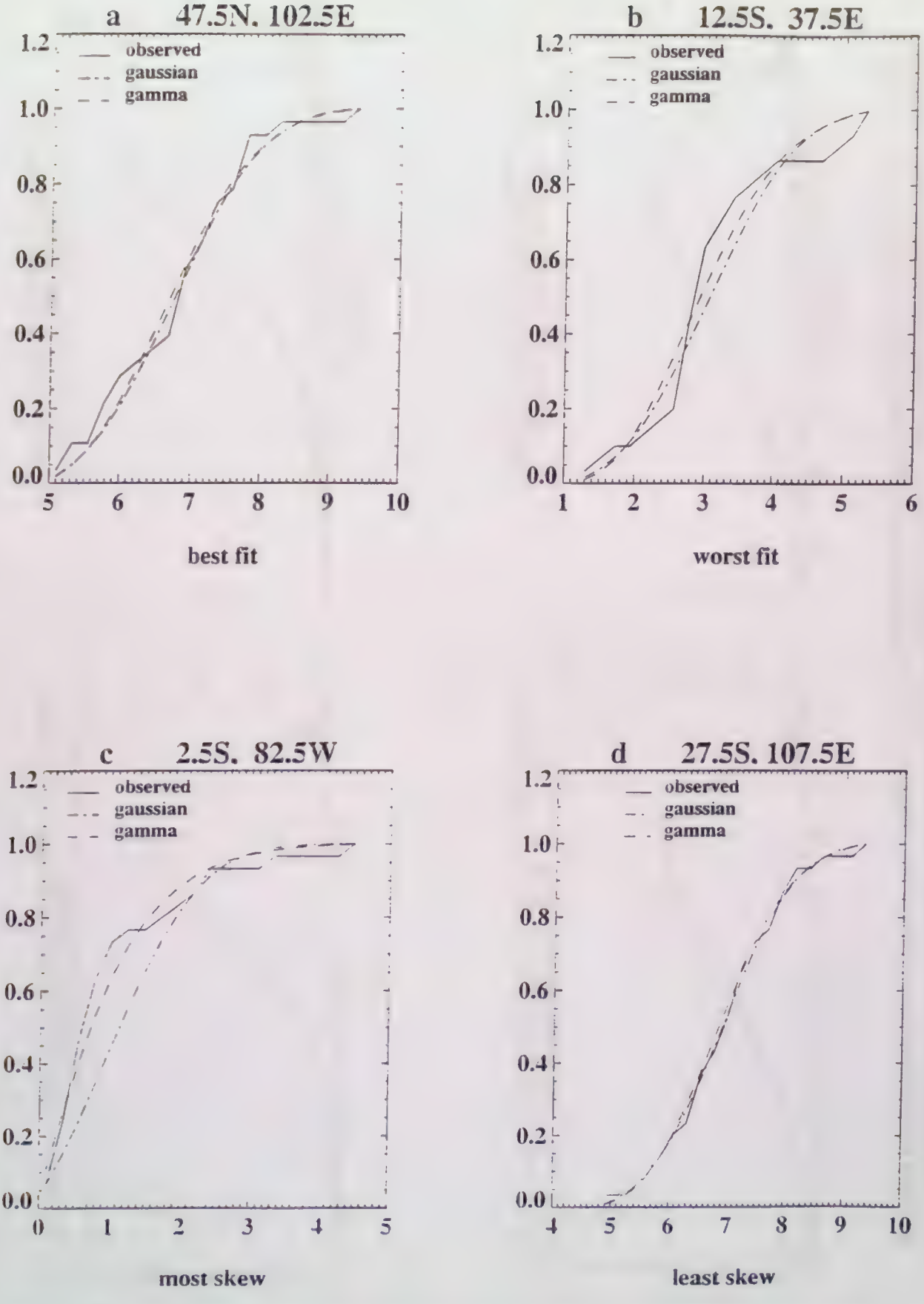


Figure 3. Examples of fitting gamma and Gaussian distributions to annual temperature anomalies.

a, The distribution best fitted by a gamma cumulative frequency distribution.

b, The distribution worst fitted by a gamma cumulative frequency distribution.

c, The most skew observed distribution.

d. The least skew observed distribution

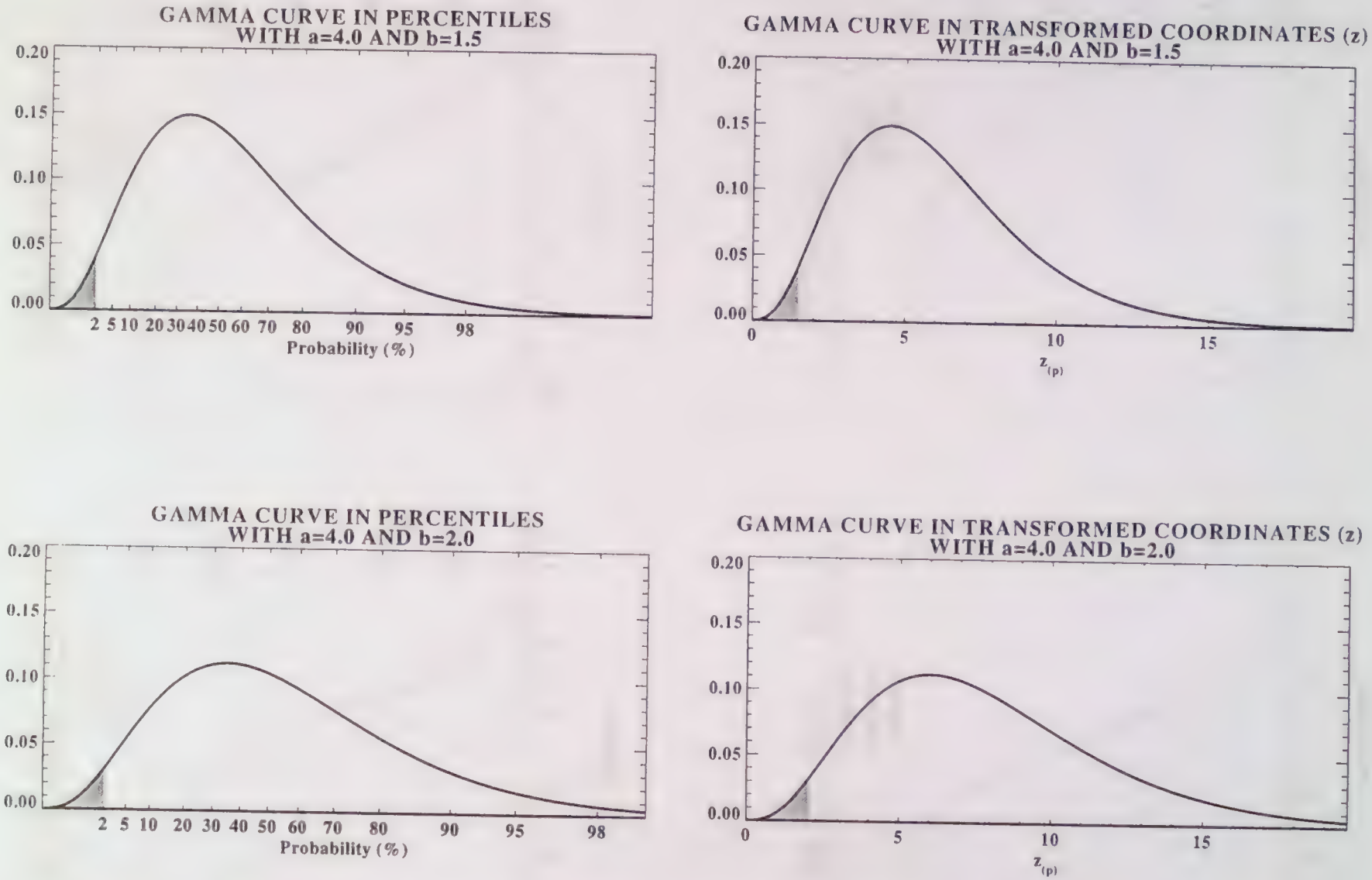
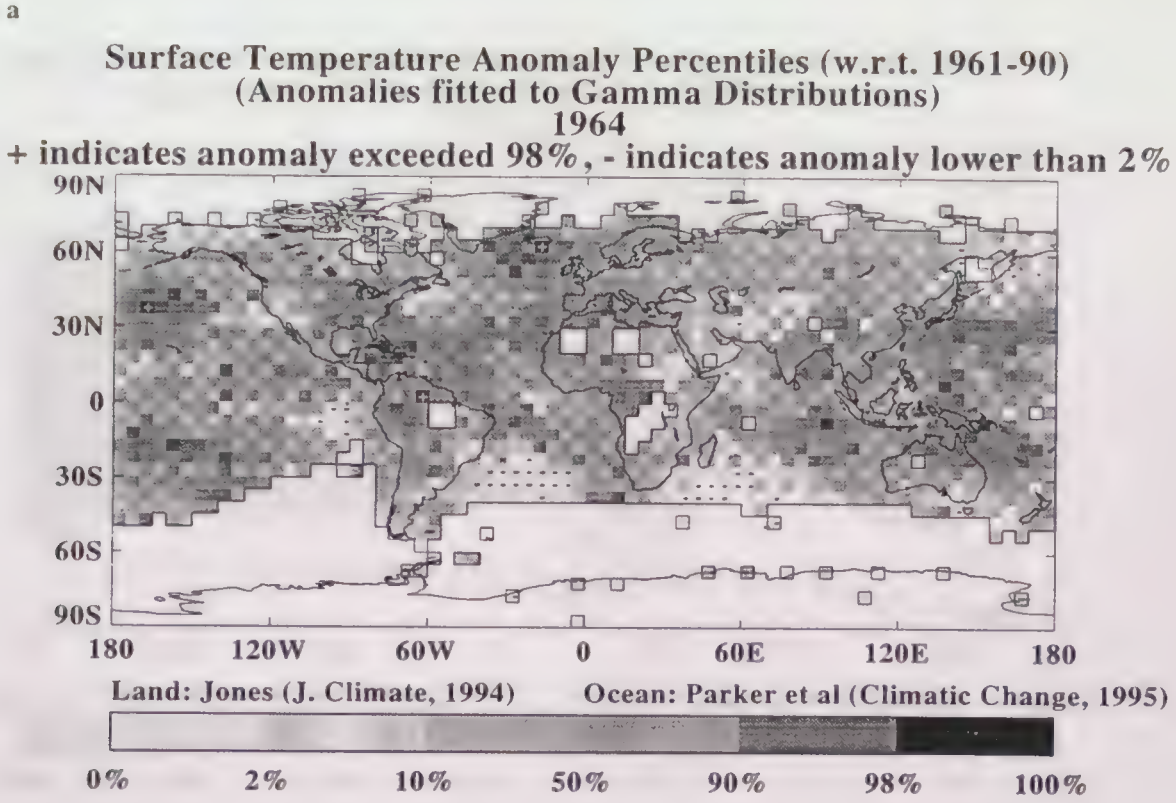
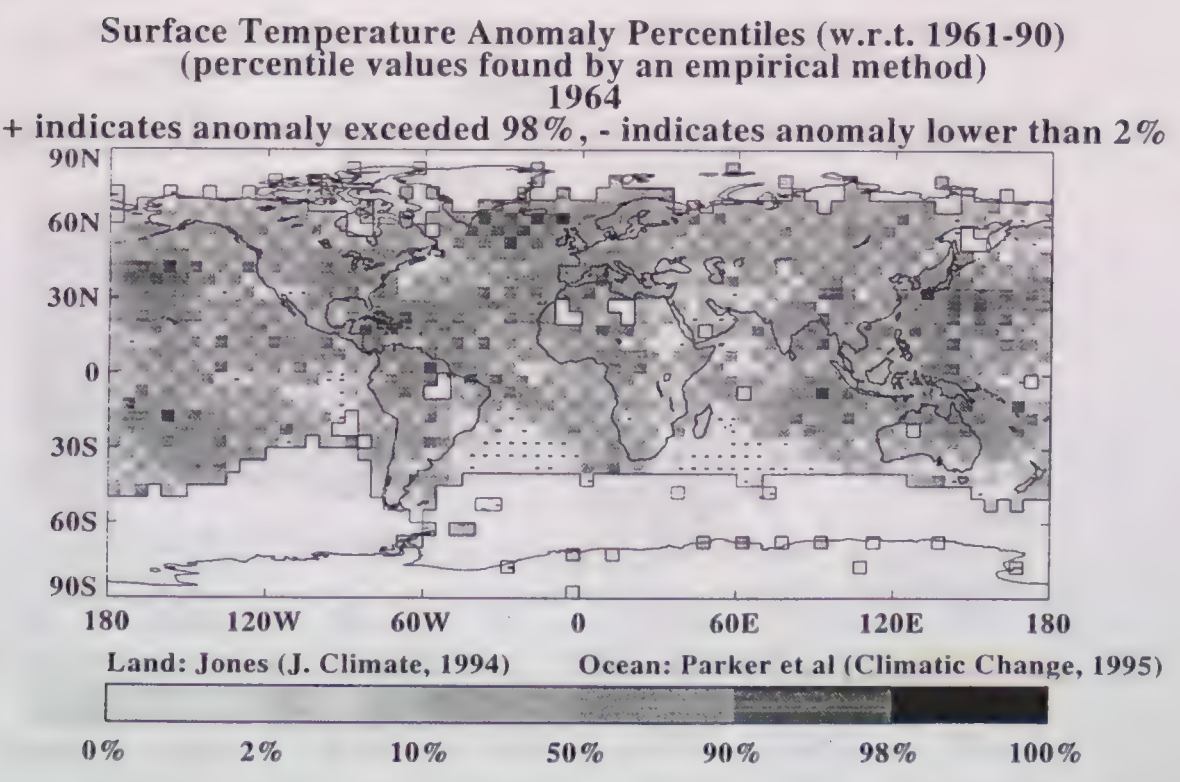


Figure 4. The influence of b on the value of the transformed gamma coordinate $z_{(p')}=b\,z_{(p)}$ .corresponding to a particular percentile. The left and right hand graphs are identical, but the abscissa has been marked in percentile points on the left graph and as actual values of $z_{(p')}$ on the right. The top graphs have b=1.5, the bottom graphs have b=2. A random value has 2% chance of being in the shaded region.

a

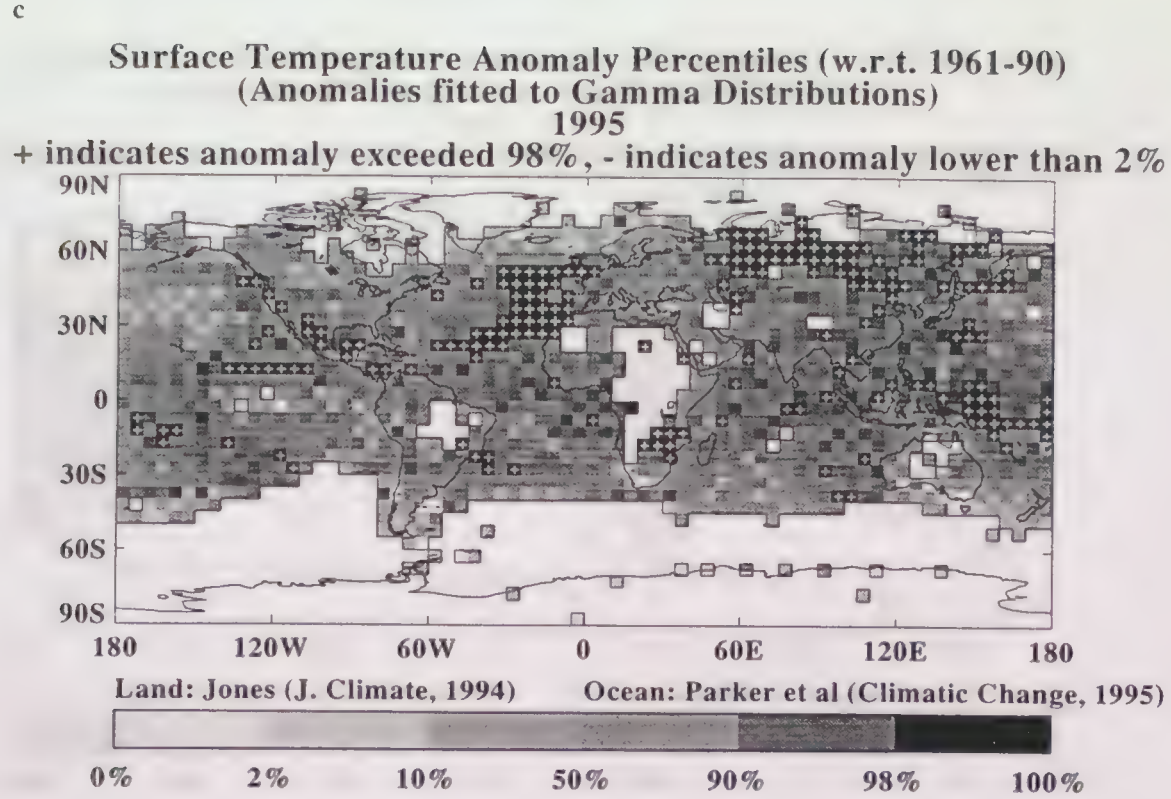


b



Figures 5a,b. Temperature anomalies for 1964, expressed as percentiles of 1961-90 data. Percentile values were calculated in (a, top) by fitting the observations to the gamma distribution and in (b, bottom) by the empirical method of Chegodayev (1953)..

c



d

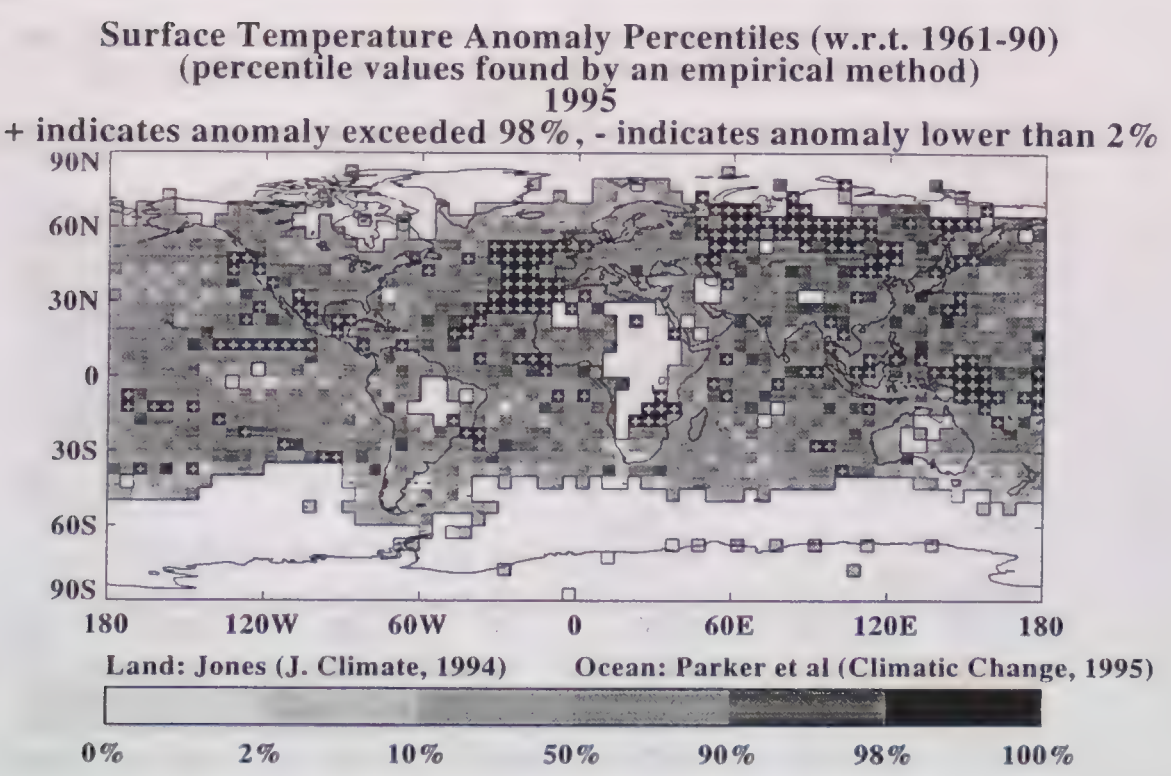
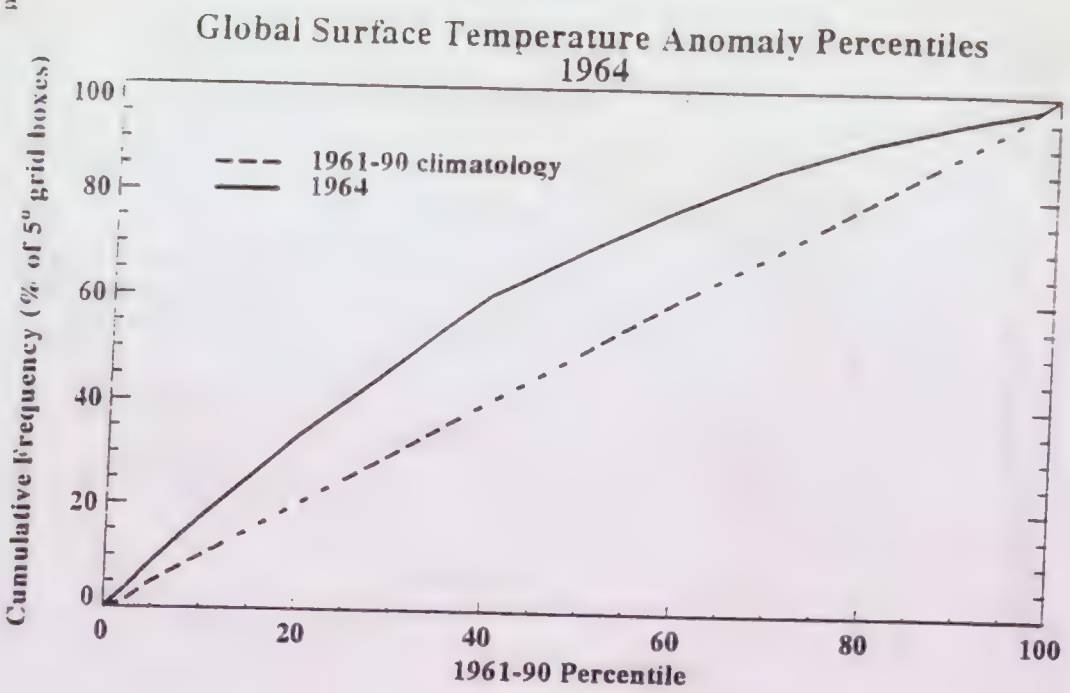


Figure 5c,d. Temperature anomalies for 1995, expressed as percentiles of 1961-90 data. Percentile values were calculated in (c, top) by fitting the observations to the gamma distribution and in (d, bottom) by the empirical method of Chegodayev (1953)

a



b

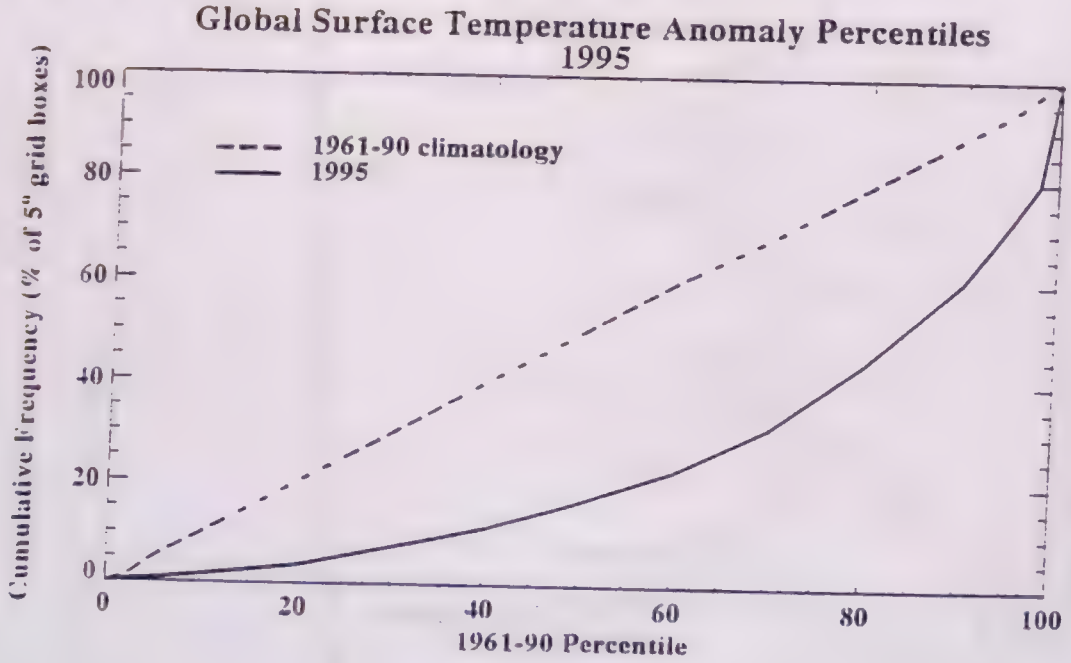


Figure 6. Percentages of anomalies which did not exceed the specified anomaly at each percentile during a. 1964. b.1995. Each of these cumulative distributions (solid lines) is compared with that for all years from 1961 to 1990 (dashed lines).

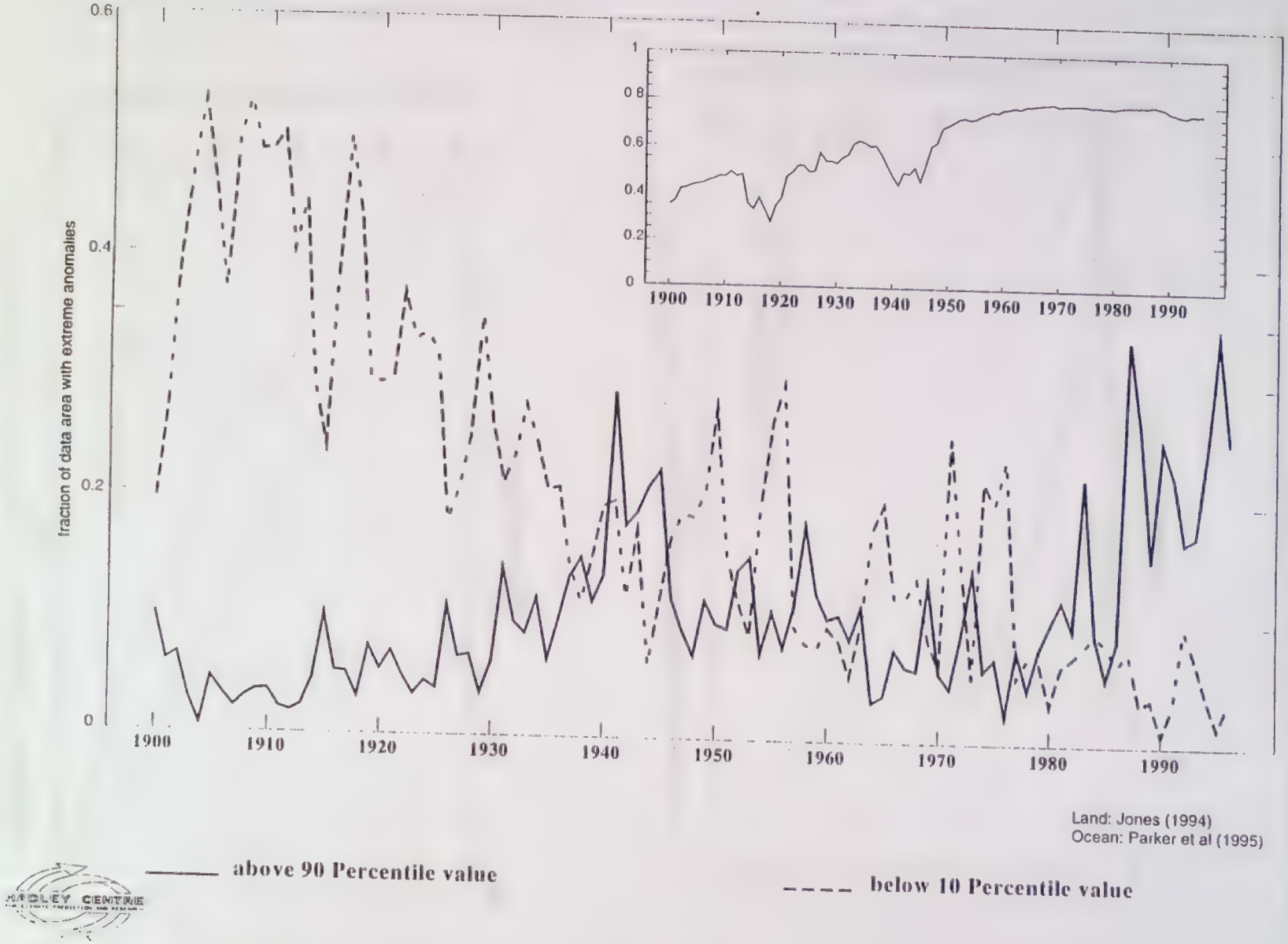


Figure 7. Percentages of the global data area with annual surface temperatures above the 90th percentile (solid line) or below the 10th percentile (dotted line). Also shown, in the inset, is the percentage of global surface area with data. The area with temperature anomalies exceeding the 90th percentile has increased by 0.11% per year, significant at the 95% confidence level. The area with temperature anomalies not exceeding the 10th percentile has decreased by 0.40% per year, also significant at the 95% confidence level.

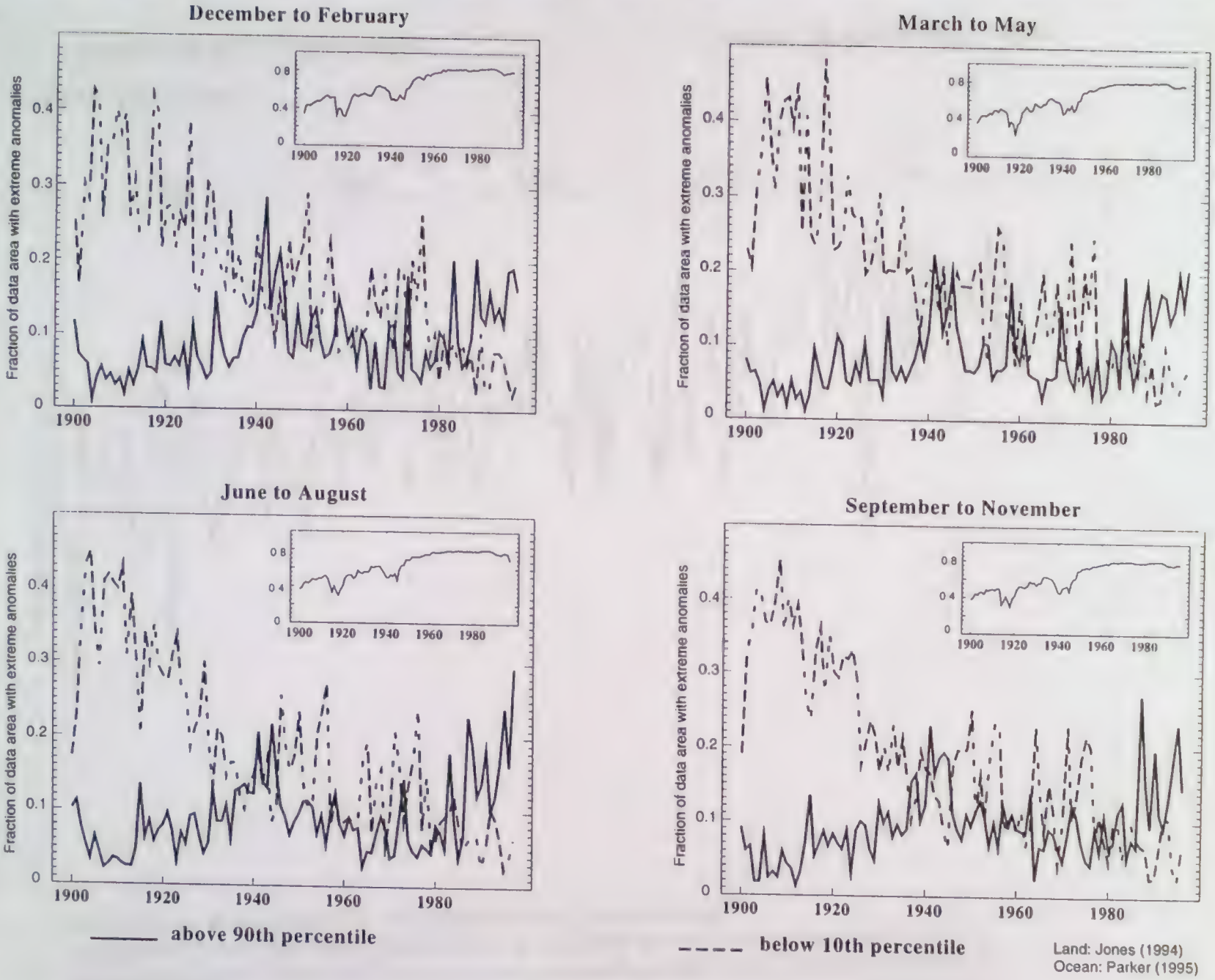
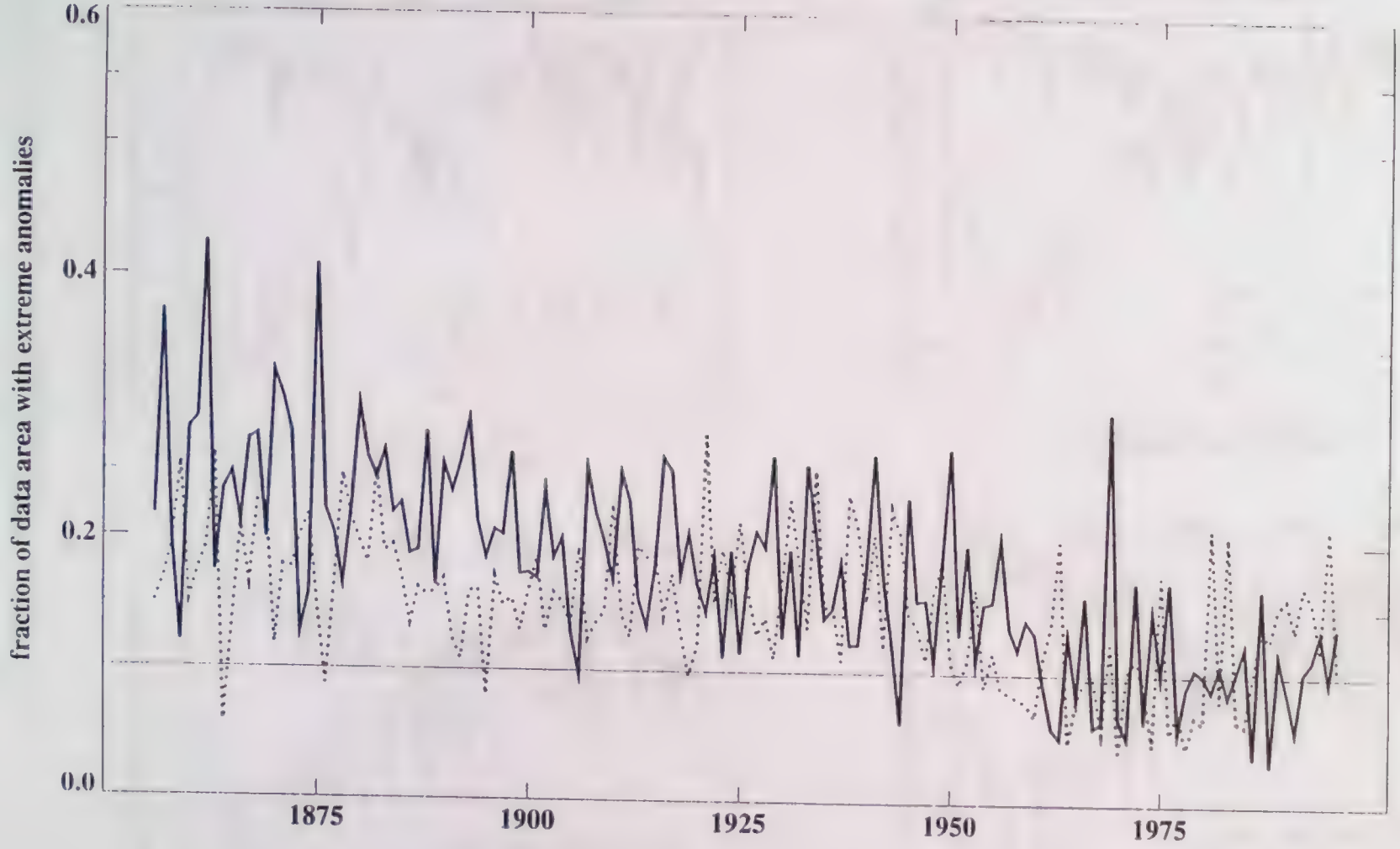


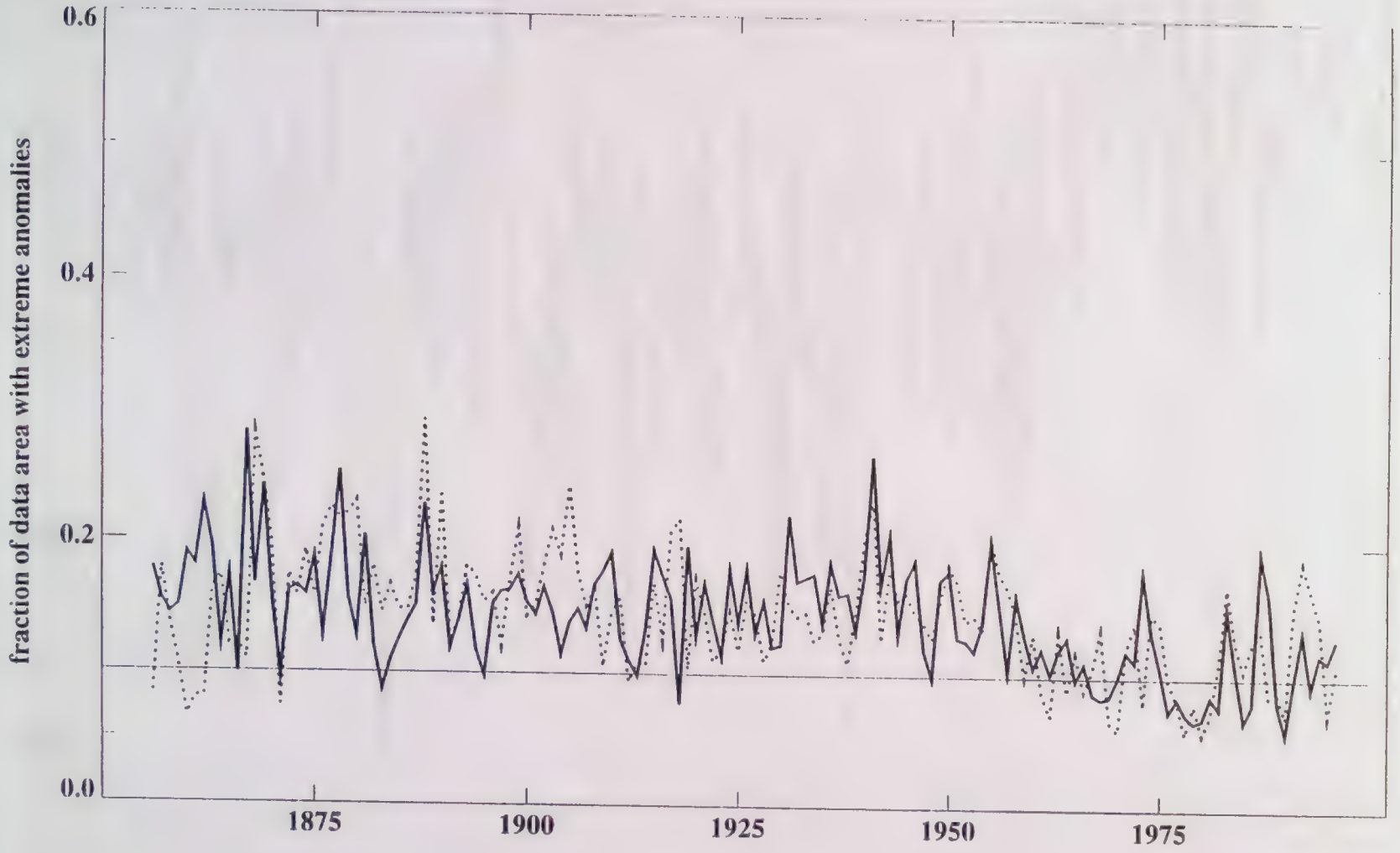
Figure 8. Percentages of global data area with seasonal surface temperatures above the 90th percentile (solid line) or below the 10th percentile (dashed line). Also shown, in the insets, are the percentages of global surface area with data in each season

Figure 9a. Fraction of global **land** data area with **adjusted** (see text) annual surface air temperatures above the 90th percentile (solid line) or below the 10th percentile (dotted line). Gridbox anomalies were first adjusted by subtracting the annual globally-averaged land air temperature anomaly in the relevant year. The percentiles were also calculated from 1961-1990 data treated in this way, so as to remove the effects of all interannual and longer-term variations of the global average.



HADLEY CENTRE

Figure 9b. Fraction of global **ocean** data area with **adjusted** annual sea surface temperatures above the 90th percentile (solid line) or below the 10th percentile (dotted line). Gridbox anomalies were first adjusted by subtracting the annual globally-averaged sea surface temperature anomaly in the relevant year. The percentiles were also calculated from 1961-1990 data treated in this way.



Ocean Parker et al (1995)

—— above 90 Percentile value

..... below 10 Percentile value

EY CENTRE

Figure 10a. Fraction of days with Central England temperature not exceeding the 1961-1990 5th percentile, 1772-1996.

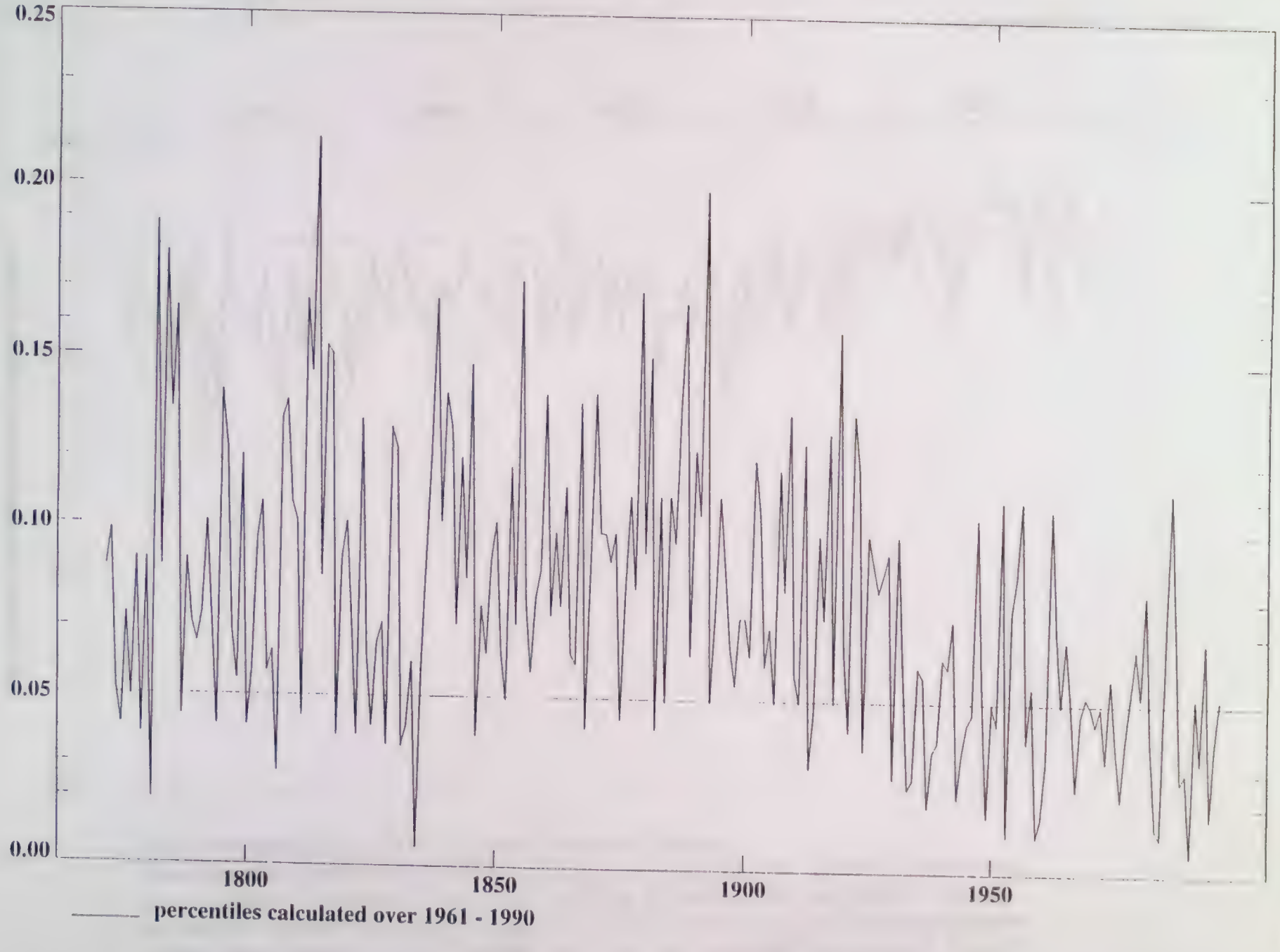


Figure 10b. As Figure 10a, but percentiles and numbers of extreme days were calculated using daily anomalies which had been adjusted by subtracting the annual mean anomaly.

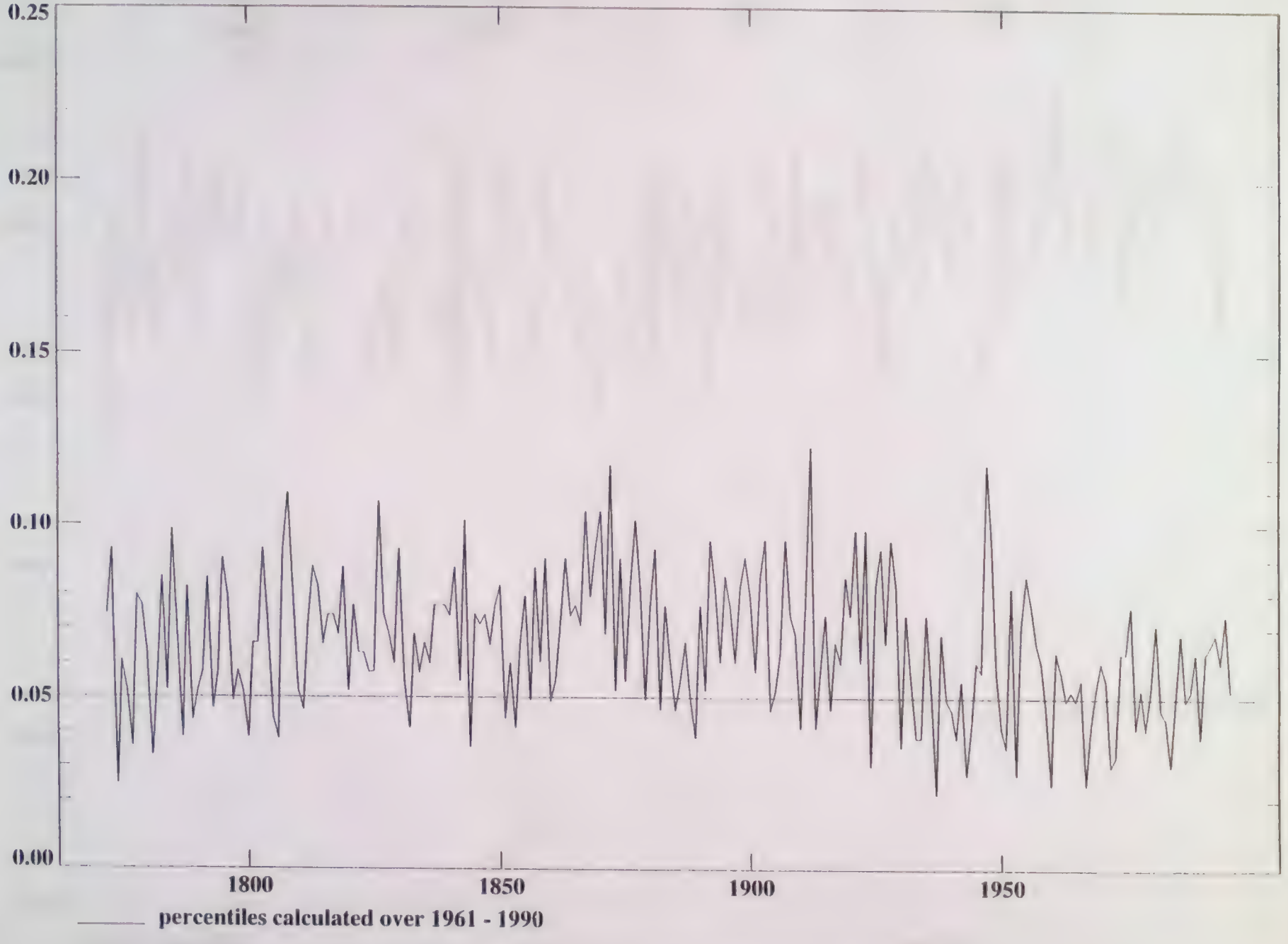


Figure 10c: As Figure 10a, but percentiles and numbers of extreme days were calculated using daily anomalies which had been adjusted by subtracting the monthly mean anomaly.

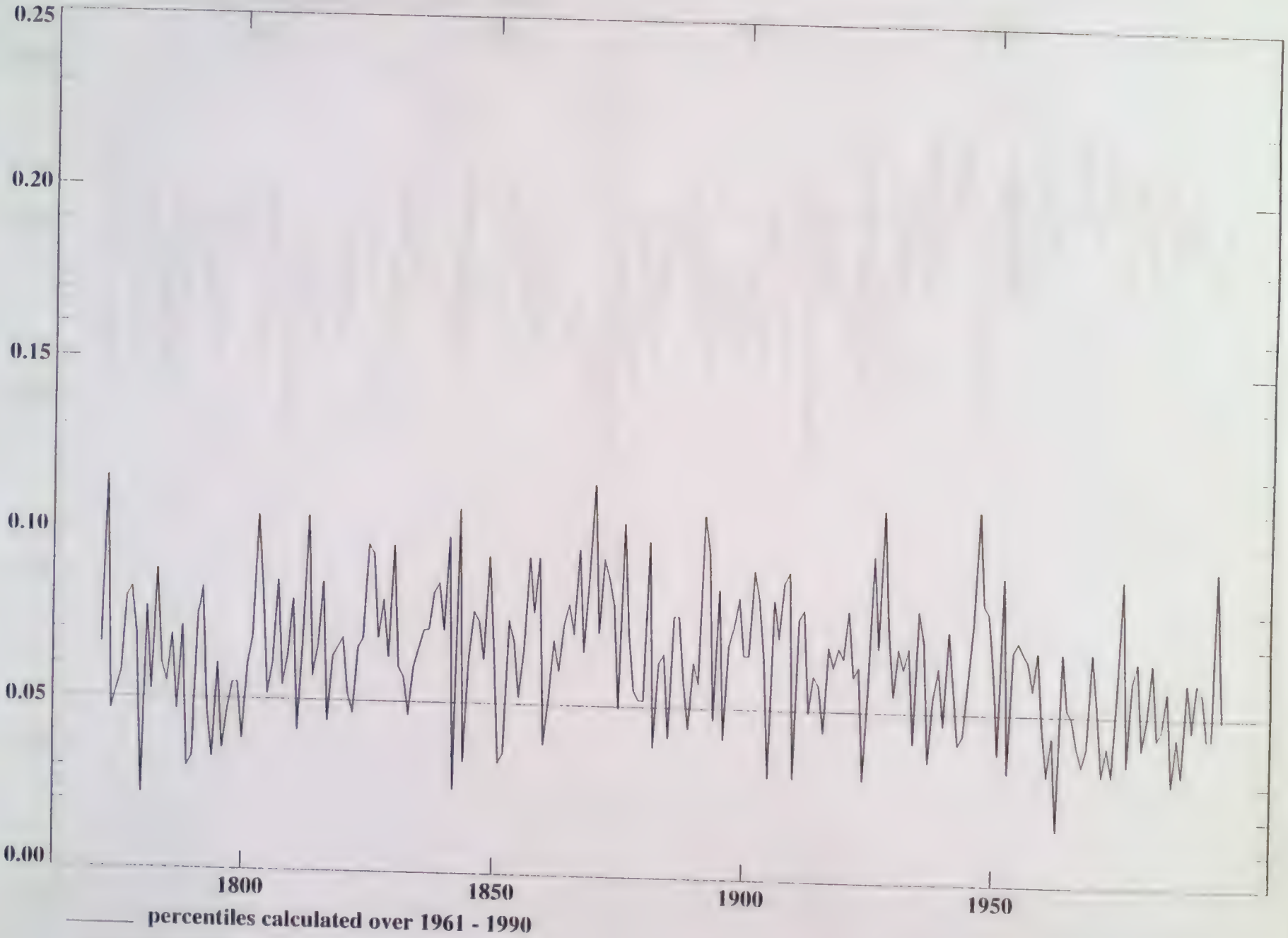


Figure 10d. Fraction of days with Central England temperature exceeding the 1961-1990 95th percentile, 1772-1996.

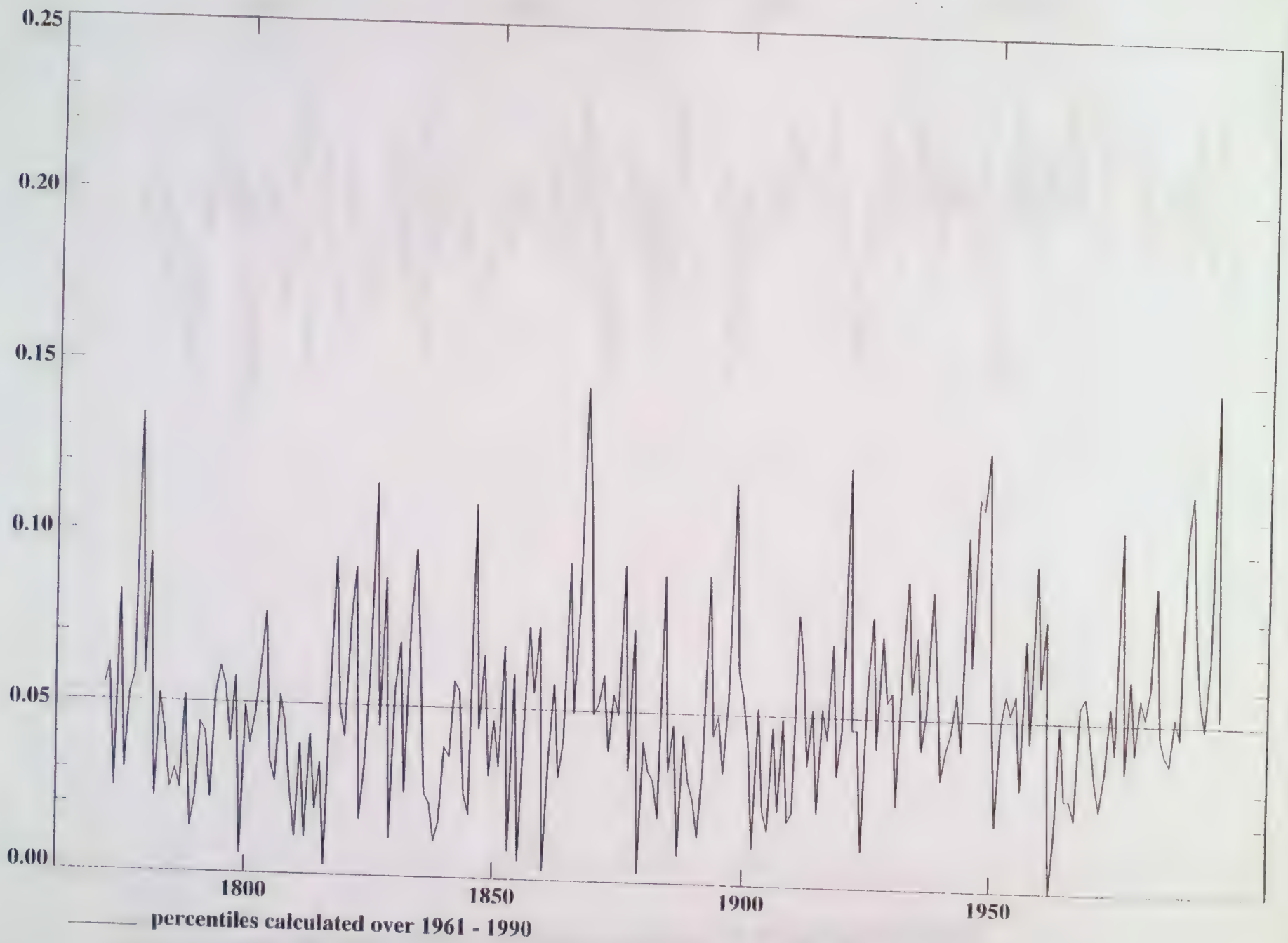


Figure 10e. As Figure 10d, but percentiles and numbers of extreme days were calculated using daily anomalies which had been adjusted by subtracting the annual mean anomaly.

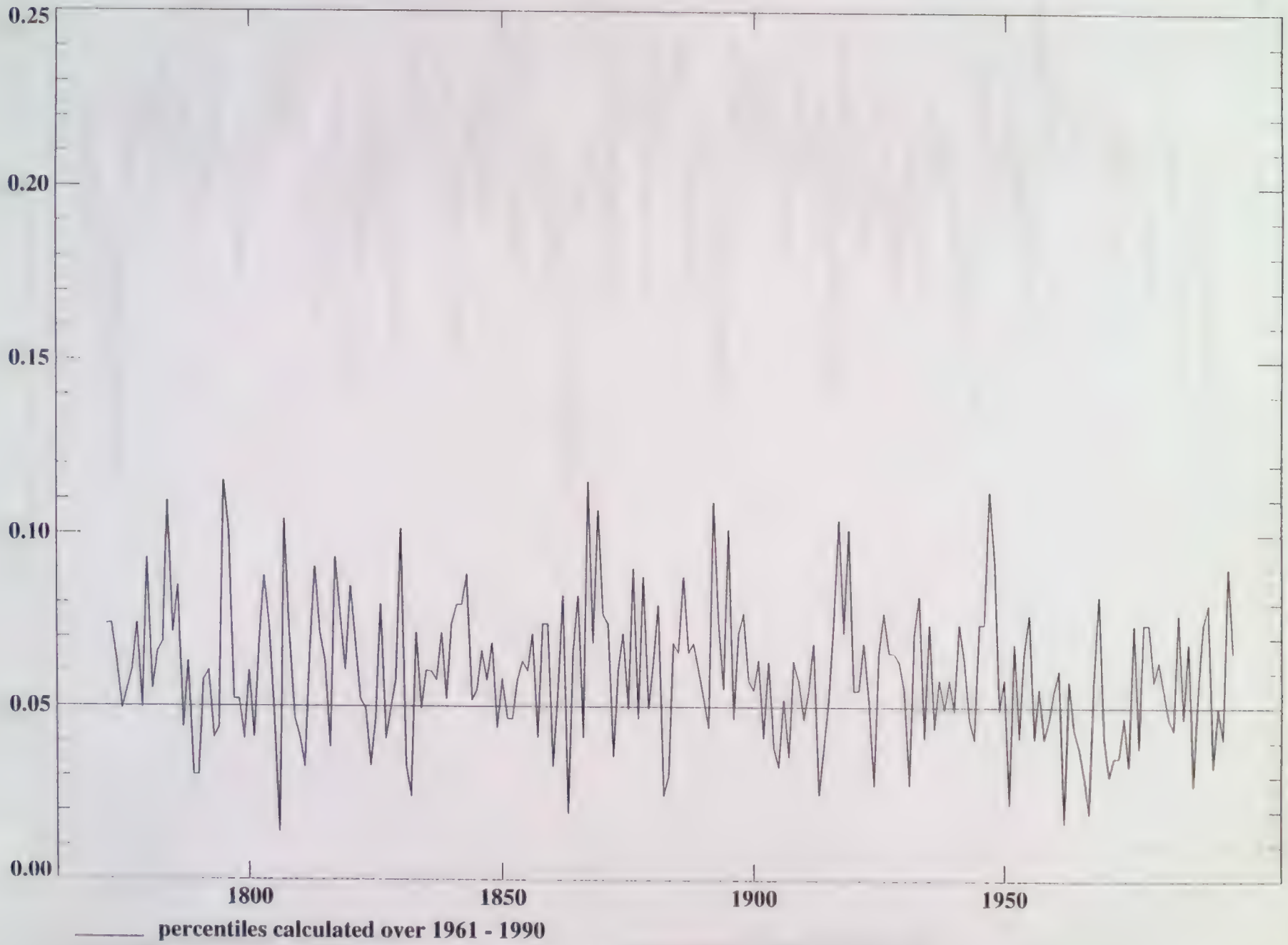
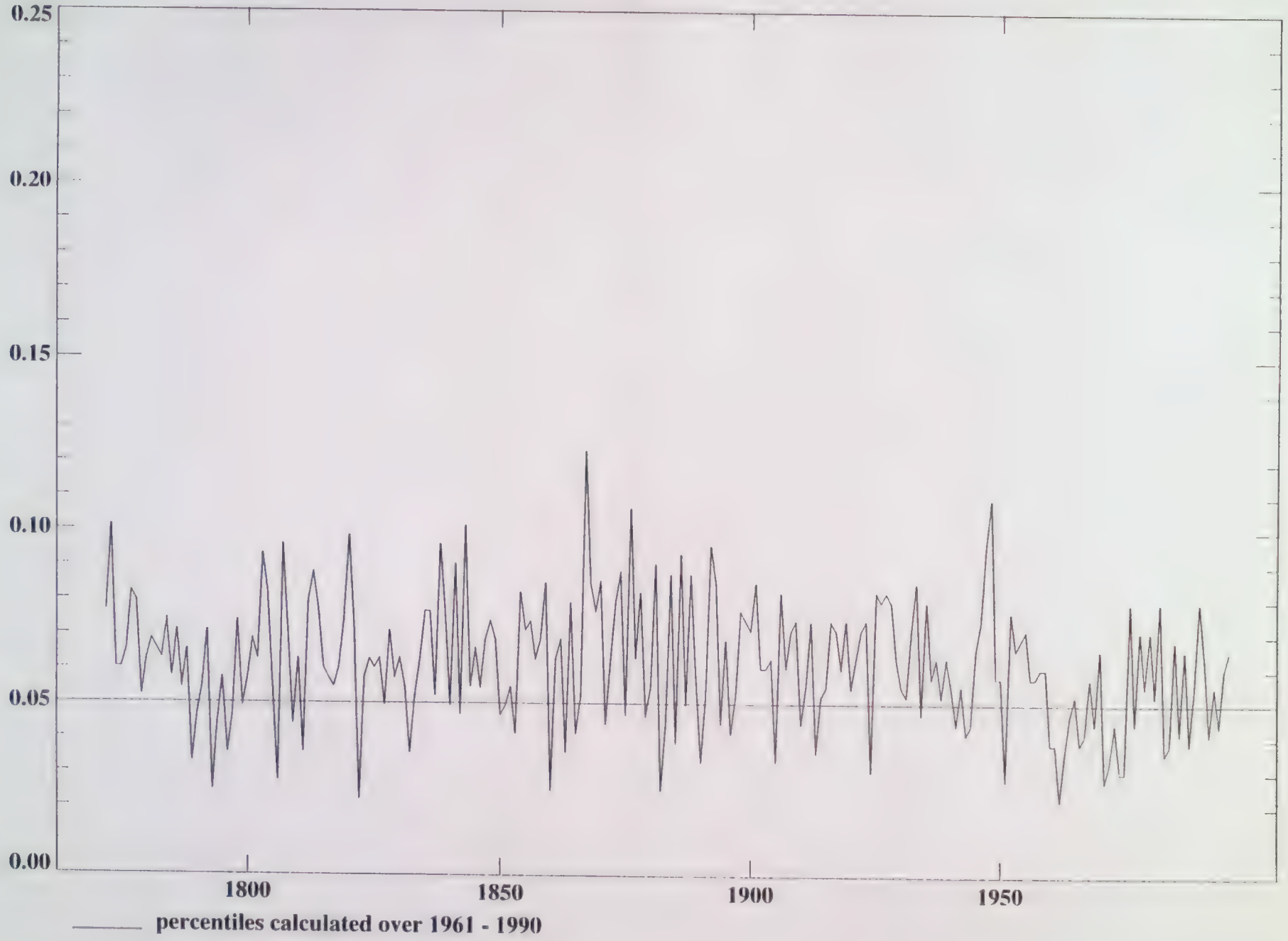


Figure 10f. As Figure 10d, but percentiles and numbers of extreme days were calculated using daily anomalies which had been adjusted by subtracting the monthly mean anomaly.



CLIMATE RESEARCH TECHNICAL NOTES

CRTN 1	Oct 1990	Estimates of the sensitivity of climate to vegetation changes using the Penman-Monteith equation. P R Rowntree
CRTN 2	Oct 1990	An ocean general circulation model of the Indian Ocean for hindcasting studies. D J Carrington
CRTN 3	Oct 1990	Simulation of the tropical diurnal cycle in a climate model. D P Rowell
CRTN 4	Oct 1990	Low frequency variability of the oceans. C K Folland, A Colman, D E Parker and A Bevan
CRTN 5	Dec 1990	A comparison of 11-level General Circulation Model Simulations with observations in the East Sahel. K Maskell
CRTN 6	Dec 1990	Climate Change Prediction. J F B Mitchell and Qing-cun Zeng
CRTN 7	Jan 1991	Deforestation of Amazonia - modelling the effects of albedo change. M F Mylne and P R Rowntree
CRTN 8	Jan 1991	The role of observations in climate prediction and research. D J Carson
CRTN 9	Mar 1991	The greenhouse effect and its likely consequences for climate change. D J Carson
CRTN 10	Apr 1991	Use of wind stresses from operational N.W.P. models to force an O.G.C.M. of the Indian Ocean. D J Carrington
CRTN 11	Jun 1991	A new daily Central England Temperature series, 1772-1991. D E Parker, T P Legg and C K Folland
CRTN 12	Jul 1991	Causes and predictability of Sahel rainfall variability. D P Rowell, C K Folland, K Maskell, J A Owen, M N Ward

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 13	Jul 1991	Modelling changes in climate due to enhanced CO2, the role of atmospheric dynamics, cloud and moisture. C A Senior, J F B Mitchell, H Le Treut and Z-X Li
CRTN 14	Sep 1991	Sea temperature bucket models used to correct historical SST data in the Meteorological Office. C K Folland
CRTN 15	Aug 1991	Modelling climate change, and some potential effects on agriculture in the U.K. P R Rowntree, B A Callander and J Cochrane
CRTN 16	Aug 1991	The Boreal Forests and Climate G Thomas and P R Rowntree
CRTN 17	Aug 1991	Development of a Stratosphere-Troposphere Data Assimilation System. R Swinbank.
CRTN 18	Sep 1991	A study of asynchronous coupling using a simple climate model. M K Davey.
CRTN 19	Sep 1991	The Oceanic Carbon Cycle. N K Taylor
CRTN 20	Nov 1991	Worldwide ocean-atmosphere surface fields in Sahel wet and dry years using provisionally corrected surface wind data. M N Ward
CRTN 21	Dec 1991	Coupled tropical ocean global atmosphere models at the UKMO. M K Davey, C Gordon, S Ineson and S Lawrence
CRTN 22	Dec 1991	Empirical parameterisation of tropical ocean-atmosphere coupling. M Allen, M K Davey, D L T Anderson and P D Killworth
CRTN 23	Feb 1992	The temporal evolution of Equatorial Currents in the Indian Ocean D L T Anderson, D J Carrington, R A Corry and C Gordon
CRTN 24	Feb 1992	Stratospheric analyses provided by the U.K. Meteorological Office M J Bailey, A O'Neill and V D Pope

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 25	Feb 1992	Modelling interannual variability in the Indian Ocean using momentum fluxes from the UKMO and ECMWF operational weather analyses. D L T Anderson and D J Carrington
CRTN 26	Mar 1992	A GCM simulation of the impact of Amazonian deforestation on climate using an improved canopy representation J Lean and P R Rowntree
CRTN 27	Mar 1992	The parameterization of rainfall interception in GCMs A J Dolman and D Gregory
CRTN 28	Jun 1992	Development of worldwide marine data eigenvectors since 1985 A W Colman
CRTN 29	Jun 1992	A tropical ocean model with reduced physics M K Davey
CRTN 30	Jun 1992	International Temperature Workshop, Boulder, Colorado, USA, 16 January 1992 Edited by D E Parker
CRTN 31	Jul 1992	Simulation of clear-sky outgoing longwave radiation over the oceans using operational analyses A Slingo and M J Webb
CRTN 32	Sep 1992	A prediction of the transient response of climate J M Murphy
CRTN 33	Nov 1992	LEPS scores for assessing climate model simulations and long-range forecasts. C K Folland
CRTN 34	Jan 1993	Stratospheric Data Assimilation System Guide Editer: R Swinbank
CRTN 35	Mar 1993	A Stratosphere-Troposphere Data Assimilation System R Swinbank and A O'Neill
CRTN 36	Mar 1993	Validation of hydrological schemes for climate models against catchment data. P R Rowntree and J Lean

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 37	Apr 1993	Modelling of palaeoclimates: Examples from the recent past J F B Mitchell
CRTN 38	May 1993	A simulation of seasonality in ENSO forecast skill M K Davey, D L T Anderson and S Lawrence
CRTN 39	Jul 1993	ENSO Prediction experiments using a simple ocean-atmosphere model D-H Wu, D L T Anderson and M K Davey
CRTN 40	Sep 1993	Diagnosis of dynamic sea-surface and sea-level changes from the Cox ocean model J M Gregory
CRTN 41	Sep 1993	Seasonal variations of the clear-sky greenhouse effect: the role of changes in atmospheric temperatures and humidities M J Webb, A Slingo and G L Stephens
CRTN 42	Nov 1993	Sea-level changes under increasing atmospheric CO_2 in a transient coupled ocean-atmosphere GCM experiment. J M Gregory
CRTN 43	Dec 1993	Global and regional patterns of climate change: recent predictions for the Arctic P R Rowntree
CRTN 44	Feb 1994	The effect of changing horizontal diffusion in the atmospheric version of the unified climate model C D Hall and R A Stratton
CRTN 45	Mar 1994	Simulation of El-Niño/Southern Oscillation like variability in a global AOGCM and its response to CO_2 increase. S F B Tett
CRTN 46	Apr 1994	Global data required for monitoring climate change. D E Parker and C K Folland
CRTN 47	May 1994	Seasonal uptake of anthropogenic CO_2 in an ocean general circulation model. N K Taylor

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 48	Jun 1994	A tropic-wide oscillation of boreal summer rainfall and patterns of sea-surface temperature. M N Ward, K Maskell, C K Folland, D P Rowell and R Washington
CRTN 49	Jun 1994	Simulation of the tropical Pacific using a simplified ocean model M A Balmaseda, D L T Anderson and M K Davey
CRTN 50	Sep 1994	Simulation of global mean temperature using a box-diffusion climate model P R Rowntree
CRTN 51	Sep 1994	Seasonal dependence of ENSO prediction skill M A Balmaseda, M K Davey and D L T Anderson
CRTN 52	Nov 1994	Simulation of the Indian Monsoon and Tropical Intraseasonal Variability by a General Circulation Model P M Inness and D Gregory
CRTN 53	Mar 1995	A comparison of modelled surface fluxes with climatological estimates. C Gordon and D K Wright
CRTN 54	Mar 1995	The representation of moist convection in atmospheric models. D Gregory
CRTN 55	Mar 1995	A consistent treatment of the evaporation of rain and snow for use in large-scale models D Gregory
CRTN 56	Apr 1995	Workshop on simulations of the Climate of the Twentieth Century using GISST, 28-30 November 1994, Hadley Centre, Bracknell, UK Edited by C K Folland and D P Rowell
CRTN 57	May 1995	Enhanced shortwave cloud radiative forcing due to anthropogenic aerosols. S E Swartz and A Slingo
CRTN 58	May 1995	The simulation of the tropical oceans in models of different horizontal resolution. C Gordon, D K Wright and C M Roberts
CRTN 59	Sep 1995	The water budget of middle latitude continental regions - a modelling and observational study P R Rowntree
CRTN 60	Sep 1995	Intraseasonal variability of the Indian summer monsoon simulated by the Hadley Centre climate model K Ashok, D Gregory and P M Inness

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 61	Sep 1995	Climate simulations with the Unified Model: AMIP runs C D Hall, R A Stratton and M L Gallani
CRTN 62	Nov 1995	Validation of surface parameters over the oceans in climate simulations with the unified model. C D Hall
CRTN 63	Dec 1995	The GISST2.2 sea surface temperature and sea-ice climatology D E Parker, M Jackson and E B Horton
CRTN 64	Dec 1995	Parametrization of momentum transport by convection. II: Tests in single column and general circulation models. D Gregory, R Kershaw and P M Inness
CRTN 65	Jan 1996	Understanding the sensitivity of a GCM simulation of Amazonian deforestation to the specification of vegetation and soil characteristics. J Lean and P R Rowntree
CRTN 66	Jan 1996	On the efficient calculation of infra-red fluxes and cooling rates using the two-stream equations. J M Edwards
CRTN 67	Jan 1996	The second Hadley Centre coupled ocean-atmosphere GCM: Model description, spinup and validation. TC Johns,RE Carnell,JF Crossley,JM Gregory, JFB Mitchell,CA Senior,SFB Tett and RA Wood
CRTN 68	Feb 1996	The sensitivity of climate simulations to the specification of mixed phase clouds D Gregory and D Morris
CRTN 69	Apr 1996	Using an ensemble of Multi-Decadal GCM simulations ro assess potential seasonal predictability D P Rowell
CRTN 70	May 1996	North Atlantic and European Seasonal Predictability using an Ensemble of Multi-Decadal AGCM Simulations J R Davies, D P Rowell and C K Folland
CRTN 71	Jun 1996	A description of the Second Hadley Centre Coupled Model (HADCM2) T C Johns
CRTN 72	Jul 1996	Palaeoclimate Modelling Intercomparison: UKMO GCM simulations for 6kBP and 21kBP C D Hewitt and J F B Mitchell

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 73	Aug 1996	ENSO responses and low frequency weather variability in the North Pacific/American sector 1949-93 A C Renshaw, D P Rowell and C K Folland
CRTN 74	Sep 1996	Version 2.2 of the Global sea-Ice and Sea Surface Temperature data set, 1903-1994 N A Rayner, E B Horton, D E Parker, C K Folland and R B Hackett
CRTN 75	Sep 1996	A new gravity wave drag scheme incorporating anisotropic orography and low level wave breaking : Impact upon the climate of the UK Met Office Unified Model D Gregory, G J Shutts and J R Mitchell
CRTN 76	Nov 1996	Middle atmosphere variability in the UKMO Unified Model R Swinbank, W Lahoz, C Douglas, A Heaps, R Brugge, W Norton, A O'Neill, D Podd
CRTN 77	Dec 1996	A high resolution AMIP run using the Hadley Centre model HadAM2b R A Stratton
CRTN 78	Jan 1997	The climate response to CO_2 of the Hadley Centre coupled AOGCM with and without flux adjustment J M Gregory and J F B Mitchell
CRTN 79	May 1997	Development of the Global Mean Sea Level Pressure Data Set GMSLP2 T A Basnett and D E Parker
CRTN 80	Jun 1997	Seasonal Forecasting D J Carson
CRTN 81	Feb 1998	Large scale modes of ocean surface temperature since the late nineteenth century C K Folland, D E Parker, A W Colman and R Washington
CRTN 82	Mar 1998	The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. P M Cox, R A Betts, C B Bunton, R L H Essery, P R Rowntree and J Smith
CRTN 83	Mar 1998	The incidence of extremes in worldwide and Central England temperatures. E B Horton, C K Folland and D E Parker